UNIVERSIDAD AUTONOMA DEL ESTADO DE MEXICO

FACULTAD DE ODONTOLOGIA

BIOESTADÍSTICA

Para la Unidad de Aprendizaje

Bioestadística

LICENCIATURA DE CIRUJANO DENTISTA

Dra. María Elena V. Escalona Franco

Septiembre, 2015

Contenido

El porqué de estos apuntes

Cuando tuve la oportunidad de iniciar la docencia en la Universidad Autónoma del Estado de México, al escuchar una serie de cifras y valores que se expresaban, en torno a las diferentes materias y temas que se iban sustentando, volaba mi imaginación pensando, como haría, posteriormente, y qué hacer para apoyar, de alguna manera a mis alumnos. Esta es la razón por la cual se elaboran estos apuntes de bioestadística con la esperanza de que pudiera aportar algún apoyo a quien tuviera interés en consultarlo.

Se supone que con la experiencia de varios años de docencia en Estadística, con la formación en Salud Publica, sea capaz de cumplir esta intención.

Estructura del documento

Después de la anterior introducción, planteo algunas consideraciones sobre lo siguiente:

- Investigación
- La aventura del manejo estadístico
- Teoría estadística
- Método estadístico
    - Conceptos
    - Formulas
    - Ejemplos
    - Interpretación

Investigación

La estadística no es más que una valiosa e indispensable herramienta para lograr, al final, un adecuado análisis de los datos que se hayan obtenido, para lograr el propósito de comprobar la hipótesis establecida y el cumplir con los objetivos de la misma.

Sin embargo, es indispensable que, tanto la lectura previa de otros documentos, o en el mismo proceso, se tenga la suficiente información para bien interpretar y saber que significan los diferentes términos o valores que se obtengan: así, cuando lea, por ejemplo, que tiene un α= 0.05: interprete que quiere decir que el trabajo o el valor, en cuestión, tiene un nivel de significancia del 5%. Habrá muchos términos que necesitaremos saber su significado, tales como: Coeficiente de variación, varianza, desviación estándar, correlación, etc.: como se lee un cuadro o una gráfica, etc. esto es lo que se pretende con este documento.

La aventura del manejo estadístico

Si en alguna etapa de nuestra vida, tuvimos la fortuna de que algún maestro de matemáticas nos lograra transmitir (o nosotros entender) el manejo de los números, nos causó tanta satisfacción, que se convirtió, el resolver los problemas o mecanizaciones en un juego, del cual nos queríamos despegar. Esto mismo debemos encontrar en la estadística al darnos cuenta de que es una falacia su dificultad.

Es pues una aventura el descubrir, por ejemplo, que cuando encontramos un valor como media (promedio), en una serie, no es solo el hecho del promedio, sino que es interesante saber que atrás de ese resultado, podamos entender que tiene una estructura que nos puede ayudar a analizar con más precisión la información.

Permítaseme, en este momento (aun cuando no se entienda del todo, ya que se verá más adelante) expone un ejemplo en el cual, con los mismos valores de una característica se podrá demostrar cuan diferente se interpretan, solo al observar su estructura.

Estableceremos, para el ejemplo, que Xi = número de pacientes al mes; fi = alumnos que manifestaron la cantidad de pacientes al mes.

Así podemos decir, por ejemplo, que 3 alumnos (fi) tuvieron 3 pacientes (Xi) al mes; 1 alumno (fi), 2(Xi); 5 alumnos (fi), 3 (Xi) etc… Veamos, pues, diferentes poblaciones con diferentes estructuras en cuanto a la frecuencia, según se tomó la información de diferentes poblaciones, las cuales las vamos a identificar con letras mayúsculas, en la parte superior de cada cuadro.

A	
Xi	fi
1	3
2	3
3	3
4	3
5	3
6	3
7	3
8	3
9	3
Suma	27

B	
Xi	fi
1	1
2	1
3	1
4	1
5	23
6	0
7	0
8	0
9	0
Suma	27

C	
Xi	fi
1	0
2	0
3	0
4	0
5	23
6	1
7	1
8	1
9	1
Suma	27

D	
Xi	fi
1	1
2	1
3	1
4	1
5	19
6	1
7	1
8	1
9	1
Suma	27

E

Xi	fi
1	1
2	2
3	3
4	4
5	7
6	4
7	3
8	2
9	1
Suma	27

F	
Xi	fi
1	1
2	0
3	0
4	0
5	25
6	0
7	0
8	0
9	1
Suma	27

G	
Xi	fi
1	2
2	0
3	1
4	0
5	22
6	0
7	1
8	0
9	1
Suma	27

En el caso concreto siempre tuvimos oportunidad de observar cómo se distribuyó cada población estudiada y que, en todo caso, podemos decir algo, en cuanto a su composición. Supóngase, ahora, que no conocemos como se estructuro y nos dan solo valores, quizá dijéramos que las 7 poblaciones estudiadas, son iguales; sin embargo, observemos ahora, una tabla en donde podamos comparar los resultados:

Población	alumnos estudiados	Promedio	Desviación estándar	Coeficiente de variación
A	27	5.0	2.58	51.6
B	27	4.6	0.975	21.2
C	27	5.4	0.998	18.6
D	27	5.0	1.491	29.8
E	27	5.0	1.92	38.5
F	27	4.7	1.129	24.0
G	27	4.7	3.65	67.6

El número de alumnos estudiados siempre fue de 27.

El promedio de pacientes fue 5 (o muy cercano) por lo que diríamos que no existe diferencia entre las poblaciones; sin embargo, observemos algunos resultados que nos apoyan en el análisis, tales como la desviación estándar o coeficiente de variación que, en términos generales, nos indican que tan diferentes o parecidas son las alumnos entre sí, dentro de cada población. Ahora podemos decir que la población C es la que más se parece entre sí, y la población G es la más diferente o dispersa, en cuanto a las diferencias de cada valor con respecto a su promedio.

El conocer la importancia que tiene el uso de la Estadística, no deberá motivar para considerarle gran valor.

Teoría Estadística

He pretendido presentar este documento, iniciando por el Concepto: Formulas, en el caso necesario; Ejemplo y la Interpretación.

Estadística

Descriptiva	Inferencial
Planeación	Asociación o correlación
Recolección	Regresión
Elaboración	Probabilidad
Análisis	Muestreo
Estadísticas básicas	Pruebas de hipótesis
	Pruebas no paramétricas

Generalidades

ESTADISTICA: es el método para recolectar, elaborar e interpretar datos.

METODO ESTADISTICO: técnica especial para el estudio de fenómenos de masa, afectadas por causas múltiples.

ESTADISTICA: METODO ESTADISTICAS: DATOS

UTILIDAD: es un indispensable auxiliar en todas y cada una de las etapas de:

Planeación	Desarrollo
Programación	Evaluación
Normalización	Investigación

Estadísticas de mayor interés

Población (censo, proyección, características, etc…)
Vitales (actos y hechos vitales).
Morbilidad (las enfermedades y sus características).
Recursos (humanos, materiales, financieros, organizativos).
Servicios (atenciones que se presentan).
Económicas
Socio-culturales
Ambientales

ESTADISTICO	CIENTIFICO

PLANEACION RECOLECCION ELABORACION ANALISIS	OBSERVACION HIPOTESIS VERIFICACION

Características del método científico

1. Único medio de manejar masas de datos numéricos.
2. Es aplicable solamente a datos que se puedan convertir en forma cuantitativa.
3. Es objetiva, los datos: sin embargo, están afectados por la necesaria interpretación subjetiva.

Planeación

- Planteamiento del problema.
- Búsqueda y evaluación de lo existente.
- Formulación de la hipótesis.
- Verificación de la hipótesis.
- Conclusiones y recomendaciones.

1. Planteamiento del problema.

Naturaleza	QUE
Importancia	POR QUE
Objetivo	PARA QUE
Solución	COMO

2. Búsqueda y evaluación de lo existente.

QUIEN	CUANDO
POR QUE	COMO
CUALES	QUE
DONDE	CUANTOS

3. Formulación de la hipótesis.

Soluciones tentativas del problema.

4. **Verificación de la hipótesis.**

- Observaciones (de qué tipo)
- Individuos: Universo, características
- Procedimientos
- Tiempo (en cuánto)
- Gastos
- Selección de personal
- Adiestramiento
- Definir técnica, unidades de medida que se observan…
- Recursos
- Balanceo de recursos: aumentar, disminuir o diferir.

5. **Conclusiones y recomendaciones.**

Notas importantes que pueden servir de guía y orientación, para el desarrollo de la investigación o el proceso de evaluación.

Proceso de investigación

- Conocimiento
- Problema
- Marco teórico (revisión documental)
- Marco conceptual
- Marco hipotético
- Variables (operacionalización)
- Instrumento (diseño o revisión)
- Muestra (tamaño y metodología)
- Aplicación del instrumento (recolección de datos)
- Ordenación de la información (elaboración de datos)
- Análisis e interpretación
- Presentación de datos
- Verificación
- Discusión (contrastar)
- Resultados
- Cuerpo (afirmaciones o negaciones)

- Presentación
- Conclusiones
- Recomendaciones
- Plantear las limitantes

Presentación de la investigación

- Titulo
- Autor (es)
- Resumen
- Introducción

    Antecedentes

    Marco

    - Teórico
    - Conceptual
    - Histórico

    Posturas teóricas (de otros y del autor)

- Método
    - Problema
    - Hipótesis
    - Consecuencias verificables
    - Identificación de variables
    - Diseño de muestra
    - Instrumento (descripción, Operacionalización)
    - Procedimientos (muestreo, instrumento, desarrollo)
    - Análisis estadístico
- Descripción del trabajo (presentación de datos)

    Interpretación de resultados

- Conclusiones, en relación a:

- La(s) hipótesis u objetivos
- La muestra y procedimientos (autocritica)
- La introducción y resultados
- Sugerencias y limitaciones

- Bibliografía
- Anexos

Guía para elaborar un programa

Quizá esta guía no sea muy propia de este manual; sino que corresponde más bien al proceso administrativo; sin embargo, considero de importancia incluir en este capítulo que se refiere a planeación.

1. Antecedentes.
2. Justificación.
    - Legal
    - Política
    - Técnica
    - Administrativa
3. Límites.
    - Espacio
    - Tiempo (duración)
    - Época (realización)
    - Universo
4. Objetivos.
    - Generales
    - Específicos
    - Mediatos
    - Inmediatos
5. Organización.
    - Estructural
    - Funcional
6. Operación.
    - Precisión de actividades
    - Planeación
    - Coordinación
    - Actividades y tareas
    - Estrategias y técnicas
    - Interna-externa
    - Cronológicas

7. Información.

- Instrumentos
- Periodicidad

8. Control.

- Supervisión (manual de supervisión)
- Evaluación (eficacia, eficiencia, efectividad, costo, etc…)

9. Implementación.

- Recursos Humanos
  Materiales
  Financieros
- Financiamiento
- Costo
- Programa
- Actividad

Recolección de los datos

Es importante que, antes de realizar cualquier proceso, hagamos la reflexión de considerar cuales podrían ser los principales errores que se pueden cometer; para que, evitándolos o previniéndolos, obtengamos mejores resultados.

Posibles errores (algunas ideas)

a) Encuestador:

Mala preparación	cansancio
Exceso de trabajo	aburrimiento
Falta de motivación	otros

b) Encuestado:

Ignorancia	tiempo inoportuno
Falta de motivación	desconfianza
Otros	

c) Unidad de medida:

Mal funcionamiento	diferentes tipos o modelos

Mala interpretación falta de acuerdo

Otros

d) Método de observación:

Falta de acuerdo falta de preparación

Falta de unificación

Reducción de errores:

- Racionalizar los conceptos Aumentar la preparación
- Vigilar las condiciones físicas seleccionar mejores técnicas
- Estandarizar métodos Investigar en similares circunstancias
- Medir errores otros

Trabajo de campo y procesamiento de la información

Es el trabajo que se realiza para desarrollar la investigación, mediante la obtención de datos.

Intramuros Extramuros

Técnicas

Obtención de datos { Observación, Experimentación, Encuesta, Entrevista } Uniformidad

Los datos deben estar justificados por los objetivos y la(s) hipótesis.

Propósitos del dato

a) Análisis

- Cualitativo y cuantitativo
- Identificación de problemas específicos
- Fundamentación de la hipótesis

b) Someter a prueba, las hipótesis establecidas

- Determinar las variables que explican el problema
- Descartar las variables poco relevantes

c) Tener elementos de juicio

- Para eliminar y corregir el problema(s) identificado
- Optimizar recursos

La decisión de la técnica está sujeta a la disponibilidad de los recursos (número, tipo, preparación, etc.)

Dependiendo del:

Indicador	**Método**	**Técnica**	**Instrumento**
Qué	Cómo	A través de qué	Con qué
	Síntesis	Sistematización	Ficha Guía
	Registro	Concentración	Cuadro
	Muestreo	Encuesta Entrevista	Cédula Guía
	Análisis	Seguimiento	Guía

Clasificación de la recolección

a) Fuente	Directa Formulario Entrevista	Indirecta Primaria Secundaria
b) Contenido	Hechos Actos Opiniones	Reacciones conocimientos
c) Forma	Libre Dirigido	Abierto Cerrado
d) Tiempo	Continua Periódica Ocasional	
e) Número	Individual	Muestreo

f) Tipos	Colectiva	Total
	Registro	Formulario Cédula Cuestionario Entrevista

Cédula

- Personalmente
- La información es homogénea
- El personal debe ser adiestrado
- Es costosa

Cuestionario

- Se envía la encuestado
- Costo reducido
- Inconvenientes
  - Mala interpretación
  - Falta de oportunidad
  - Datos o comprobables
  - Faltan explicaciones
  - Analfabetismo

Entrevista

- No se exhibe el formulario
- Mayor flexibilidad
- Se puede comprobar
- Inconvenientes:
  - No siempre se está dispuesto a informar con veracidad
  - Puede fracasar por incapacidad del encuestador

Normas de observación

Condiciones previas:

1. Antes de comenzar el trabajo sobre el terreno, el observador debe familiarizarse completamente con lso objetivos de su investigación.
2. Las técnicas de observación y de anotación deben ser ensayadas con antelación y, si es necesario den retirarse a fin de obtener notas de buena calidad, sobre el terreno.
3. Antes de comenzar una observación, el observador debe memorizar un alista de control de los elementos que se propone observar.

Procedimiento:

4. Las observaciones deben ser anotadas sobre el terreno, en la medida en que las circunstancias lo permitan; en caso contrario, lo más pronto posible.
5. El intervalo admisible entre la observación y la anotación, se mide en minutos o, en caso de condiciones particularmente difíciles, en horas. Las observaciones que se guardan en la cabeza hasta el día siguiente, deben considerarse como perdidas.
6. La relación entre el tiempo pasado en la observación y el tiempo pasado en la anotación, está en función de la naturaleza de la investigación, pero no conviene limitar el tiempo de la anotación, con objeto de obtener periodos de observación más prolongados.
7. El observador no debe olvidar que forma parte del sujeto de observación y que es necesario que anote sus propias acciones durante el periodo de observación.

Contenido:

8. Las notas deben incluir fecha, hora y duración de la observación: el lugar exacto (con mapas, fotografías y croquis, si es necesario); las circunstancias, las personas presentes y su función: la función atribuida al observador; los aparatos y el equipo utilizados; los aspectos determinantes del ambiente físico (temperatura, luminosidad, ruidos, etc.) y todas sus eventuales modificaciones.

9. Las opiniones, las hipótesis inverificables, las deducciones o las observaciones sobre el carácter o la personalidad de los sujetos, deben ser eliminados.
10. Las conversaciones y los diálogos deben ser transcritos en estilo directo. Aún cuando es posible una transcripción completa; los resúmenes deben ser anotados, en primera persona.
11. Las opiniones y las deducciones sacadas de las notas del observador deben ser anotadas separadamente en un diario de investigación o en una agenda, de manera regular.

Ordenación:

12. Las notas deben ser revisadas lo antes posible, con objeto de efectuar en ellas las correcciones y adiciones necesarias.
13. Las notas deben ser clasificadas provisionalmente, antes de la elaboración de un sistema de clasificación definitivo, indicando claramente en cada una de ellas la clarificación correspondiente.

Diseño de formularios

Información	completa
	Eficiente
Datos:	Administrativos (identificación)
	Del problema (definidos)
	a) propósito
	b) circunstancias

1. Decidir datos	Deseables Prácticos	Factibles Limitados
2. preguntas	Orden lógico	Entendibles
3. Respuestas	Espacio adecuado Pueden ser precodificadas	
4. características	Quién encuesta Dónde	A quién Cuándo

	Cuanto tiempo A cuantos	Cómo Características del sujeto
5. prueba de operabilidad y validez		
6. Instructivo	Propósito Manejo	Llenado Destino

Presentación de la información

La presentación puede ser: **TABULAR, GRAFICA, TEXTUAL, MIXTA**

TABULAR

Título: Debe responder a las preguntas: QUE, COMO, DONDE, CUANDO no olvidar que sea preciso, conciso, completo, claro.

Puede agregarse número de cuadro, cuando son varios.

EJEMPLO:

CUADRO No. 1
POBLACION, CON O SIN EXPRESION ESCOPTOFILICA,
POR GRUPOS DE EDAD Y SEXO
SAN FELIPE, CHIS.
2012

Cuadro

Columna matriz (variable independiente)

Columnas de distribución (variable dependiente o de frecuencias)

Encabezados de columnas

Regiones o líneas

Columnas

Líneas de referencia

Líneas y columnas de totales

Cuadro 1. Vegetación y usos del suelo en la Sierra Madrigal en el período 1973-2003

Vegetación y uso del suelo	Superficie por año					
	1973		1984		2003	
	ha	%	ha	%	ha	%
Relicto de selva alta perennifolia	2 627.1	72.1	1 807.8	49.6	584.4	16.0
Vegetación secundaria	471.6	12.9	1190.6	32.7	2 032.4	55.8
Vegetación caldcóla					22.5	0.6
Pastizal	503.6	13.8	616.1	16.9	885.0	24.3
Agricultura de temporal	24.7	0.7	27.3	0.7	59.1	1.6
Agricultura perenne	14.9	0.4			57.3	1.6
Asentamientos humanos					1.2	0.0
Total	3 641.9	100.0	3 641.9	100.0	3 641.9	100.0

PIE: Notas aclaratorias u observaciones

Fuente

EJEMPLO:

Posibles errores:

a) Disposición incorrecta
b) Títulos y encabezados incompletos
c) Cuadros, con solo porcentajes
d) Solo números absolutos
e) Cuadros sobrecargados

Lectura de un cuadro:

a) Lectura vertical y horizontal
b) Leer títulos y encabezados
c) Notas explicativas
d) Averiguar unidades de medida
e) Conocer las unidades de resumen (promedios, tasas.)
f) Relacionar los valores centrales con cada parcial, o grupal, o individual
g) Relacionar, entre sí, los valores de cada modalidad o variable
h) Detectar irregularidades o inconsistencias entre las mismas variables, o con experiencias anteriores
i) Leer las conclusiones del análisis del cuadro que se ha interpretado

Plan tabular

Antes de proceder al recuento, debe determinarse cuántos y cuáles serán los cuadros que deseo presentar, para dar respuesta al análisis que satisfaga a la(s) hipótesis; así como, si se presenta una o más variables en combinación o asociación.

Para esto voy a llenar un cuadro que se llamara Plan Tabular y que consiste en tener en la columna matriz la relación completa de las variables, de que consta el formulario; esta misma lista se va a colocar en las columnas de distribución.

EJEMPLO:

Supóngase que se tienen las variables de: Edad, Sexo, Religión, Grado de religiosidad, Lugar de nacimiento, Lugar de residencia, Escolaridad, Ocupación, Estado civil, Ingreso económico, Expresión comportamental de la sexualidad y Preferencia sexual (12 variables) en el formulario.

Realizo mi plan tabular para determinar los cuadros a presentar.

VARIABLES	Sexo	Religión	Grado rel	Lug nacim	Residenci	Escolarid	Ocupación	Edo civil	Ingreso	Expresión	Preferenc
Edad	X	X				X	X	X		X	X
Sexo		X	X	X		X	X	X		X	X
Religión			X	X	X	X		X		X	X
Grado de religiosidad				X	X	X		X		X	X
Lugar de nacimiento					X	X	X	X		X	X
Lugar de residencia						X	X			X	X
Escolaridad							X	X	X	X	X
Ocupación								X	X	X	X
Estado civil									X	X	X
Ingreso económico										X	X
Expresión comportament											X

Así, puedo elaborar cada cuadro. Con todas las columnas que necesito y, al ir computando, puedo ir llenando los cuadros.

Nótese que puedo determinar la combinación de la o las variables que deseo presentar, para el análisis. Se puede apreciar en el ejemplo tabular, de páginas atrás, que se realizó un cuadro en que combinaba: edad con sexo y la presencia de la expresión escoptofílica.

Es evidente que no debo abusar de presentar muchos cuadros; sino estrictamente los necesarios. Inclusive puedo hacer muchos más, que me servirán para tener algunas apreciaciones (por fuera) y solo presentar los relevantes, para no cansar al lector de la investigación. Recuérdese, por otro lado que los cuadros que se presenten, deben responder a la(s) hipótesis y objetivos de la investigación.

Textual

Es la presentación de los datos en base a solamente redacción, sin más que la explicación, enumerar cifras, etc…

Mixta

Es la presentación, combinando cualquier tipo: textual, tabular y/o gráfica.

Grafica

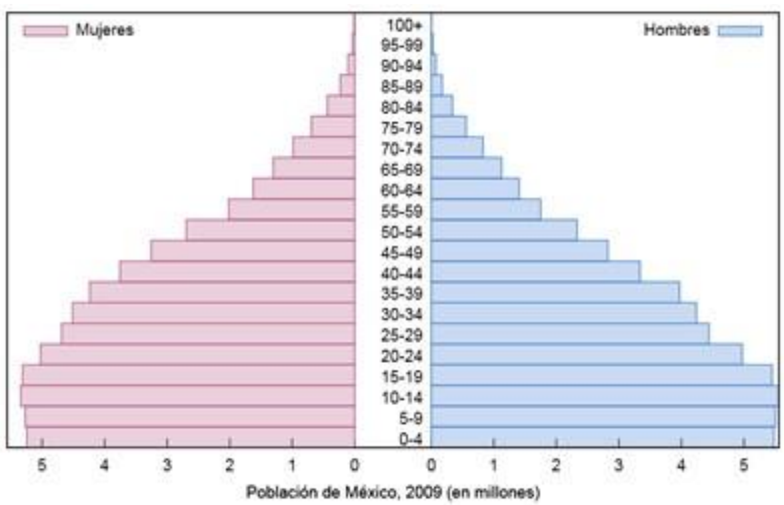
TITULO: El mismo que le hayamos puesto al cuadro

EJEMPLO:

GRAFICA No. 1

Gráfica: Dibujo para presentar la información, dependiendo de la clasificación de los datos, según su escala o serie.

EJEMPLO:



Pie:

Notas aclaratorias u observaciones
Claves de "acsurados", o colores
Fuente: cuadro….

Características:

1. Presenta objetivamente
2. Permite establecer comparaciones
3. Proporciona estimaciones o predicciones (nomograma)
4. Auxiliar en el análisis
5. Puede no ser exacta
6. Si no se usa buena técnica, se falsea
7. Puede haber cuadro sin grafica; pero no, grafica sin cuadro.
8. El titulo se coloca al centro, en la parte superior de la gráfica; si son varias, se numeran.
9. Proporción entre las coordenadas:

$$Y : 1 :: X : \sqrt{2} = Y : 1 :: X : 1.5$$

a) Variable dependiente: Y, o vertical, u ordenada (columna(s) de distribución).
Variable independiente: X, u horizontal, o abscisa (columna matriz).
b) Unidades de medida (títulos):

Variable independiente “X”, abajo y centrado.
Variable dependiente “Y”, arriba a la izquierda, en la parte superior del eje “Y”.

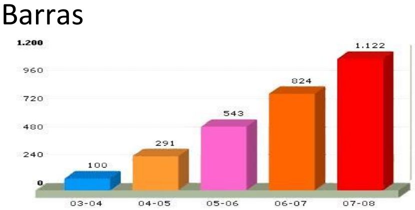
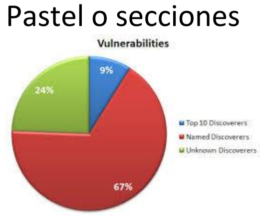
c) Puntos de referencia de unidad, a intervalos iguales de sus respectivos ejes (líneas).
d) En la variable dependiente (Y), se debe indicar con “O”, de preferencia.
e) Usando técnica adecuada se pueden cortar las barras de la gráfica.
Recuérdese que la gráfica es “objetiva”; por lo tanto, es preferible no hacerlo.

Selección de grafica según escala o serie.

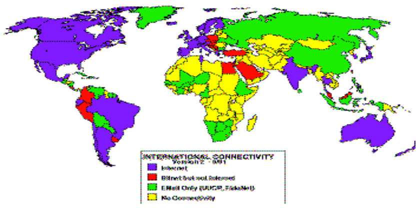
Se presentan a continuación dos guías para la seleccionar el tipo de gráfica, según su escala y/o serie. La primera es un cuadro con solo la determinación del tipo de a seleccionar; la segunda resulta más completa, en virtud de que se presenta algún ejemplo de gráfica.

Tipo de grafica		Escala			Serie		
		Cuali-tativa	Cuantitativa		Crono-lógica	Espacial	Asocia-ción
			Continua	Discreta			
Barras, bastones					???		
Pastel, circular					???	???	
Picto-grama	Carto-grama						
	Figu-ritas						
Histograma							
Polígono de frecuencia							
Curva integral							
Correlación (nubes)		???	???	???			
Línea de regresión		???	???	???			
Polar							
Semilogarítmica							

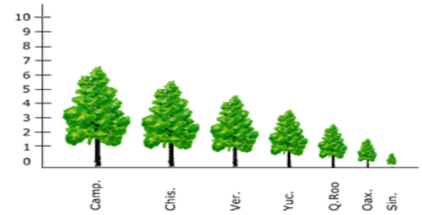
Presentacion gráfica guía para determinar la grafica

Grafica	Escala			Serie		Grafica
	cualitativa	cuantitativa		criológica	Geográfica	
		continua	discreta			
Barras 						Pastel o secciones 

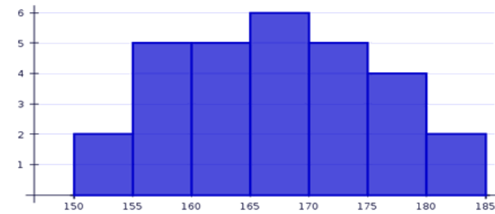
Cartograma



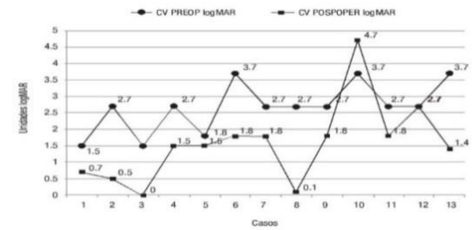
Pictograma – figuras



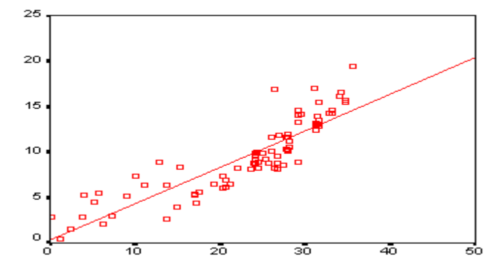
Histograma



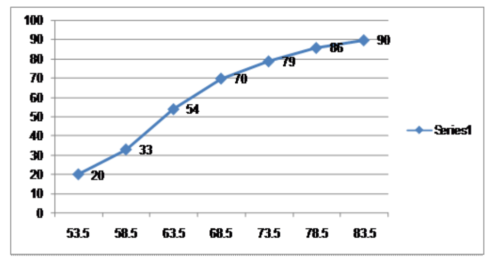
Polígono de frecuencias



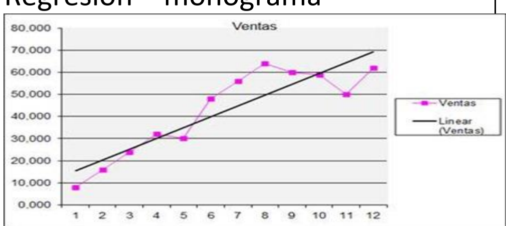
Correlación o asociación (Nubes)



Curva integral (Ojiva)

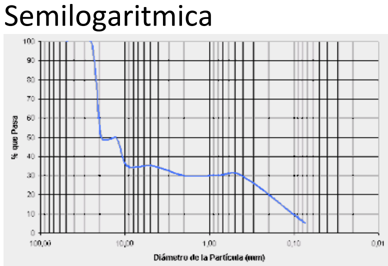


Regresión – monograma

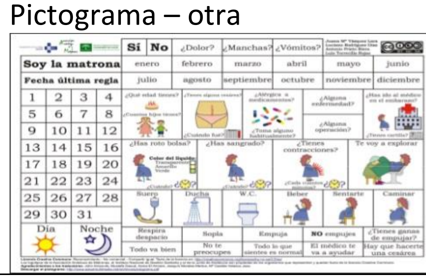


Polar

Semilogaritmica



Pictograma – otra



Análisis de la información

Significado de algunos símbolos y términos

N UNIVERSO

Número de elementos estudiados o a estudiar, en una investigación, considerando al total de la población (N).

n TAMAÑO DE LA MUESTRA

Número de elementos estudiados o a estudiar, como parte (n) de toda (N) la población.

EJEMPLO:

El Estado de México cuenta con 12 446 846 habitantes, para el 30 de junio de 1988.

Se encuestan a 12 737 habitantes.

En et caso: N= 12 446 846

n= 12 737

$\sum$ Sumatoria

Indica que es la suma de los valores observados, en los elementos del estudio.

EJEMPLO:

Pedro	17 años
Jacinto	49 años
teresa	23 años

N= 3 elementos (individuos) estudiados.

$\sum =$ 89 años, que suman entre los 3 individuos.

Indicadores y estimadores

Dentro de la estadística, ya hemos visto que se maneja la estadística descriptiva y la inferencial. Por otro lado existen infinidad de fórmulas de mayor o menor complejidad, sin embargo, deseo tratar, solamente algunas de ellas.

Descripción de datos

1. Clase
2. Límite de clase
3. Rango
4. Amplitud
5. Centro de clase

DESCRIPTIVA

6. Razón
7. Proporción
8. Porcentaje (%)
9. Prorrateo
10. Tasa
11. Media ($\bar{x}\ \mu$)
12. Mediana (Md)
13. Moda (Mo)
14. Coeficiente de verificación (CV)
15. Varianza ($S^2\ o\ \sigma^2$)
16. Desviación estándar (S o σ)
17. Cuartil (a) (O)
18. Decil (a) (D)
19. Percentil (a) (P)
20. Momento
21. Sesgo
22. Curtosis
23. Índice de correlación (I´)
24. Asociación
25. Regresión
26. Calculo de población
27. Índice endémico

INFERENCIAL

28. Estadístico o indicador
29. Parámetro
30. Estimadores
31. Tamaño de muestra (n)
32. Grado de significancia
33. Grado de libertad (v)
34. Nivel de confianza (z)
35. Error estándar (EE, σ_x)
36. Elementos del binomio (a+b)
37. Probabilidad (Pr)
38. Probabilidad. Normal
39. Probabilidad. Binomial
40. Probabilidad. Poisson
41. Pruebas de significancia
42. Hipótesis nula (Ho)
43. Hipótesis alternativa (H_1)
44. Tipos de error
45. Ji cuadrada (x^2)
46. T de Student (t)

Ahora trataré de explicar o dar interpretación del significado de cada uno de los conceptos anteriores para que cuando nos encontremos con alguno de ellos, sepamos de qué se trata. Así mismo se expresará la fórmula y algún ejemplo.

Guía para la selección de analizar

NOTA: Resultará difícil que en este momento se puede entender plenamente esta guía: es recomendable que se lea someramente este capítulo de la clasificación y determinante de indicadores, o aún que no se lea y después de revisar los conceptos de los indicadores y su aplicación se regrese a estudiar esta parte para poderse entender con mayor facilidad.

Lo más importante es haber clasificado la variable o determinando el uso que le voy a dar los datos, según el análisis que requiero.

El cuadro siguiente pretende ayudar a determinar el analizador que necesito, según la característica variable o dato. Al determinar la escala o uso del dato encontrare enseguida las diferentes alternativas a seleccionar y aplicar el analizador. Estas alternativas están señaladas como un número que corresponde a la relación de indicaciones que se enlistaron con anterioridad y que cada uno está identificado con el número que le antecede, así por ejemplo el "cuartil) tiene el número 17. Con este número, pués, identificaremos los identificadores y solamente señalaremos el o los números después de la clasificación o las posibilidades de análisis, para seleccionar justamente lo que necesitamos, pués no se tiene que analizar un dato con todos los indicadores o procesos que se enumeran.

Cualitativa	6, 7, 8, 9, 10, 24, 25, 26, 27. 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48
Cuantitativa	6, 8, 9, 11, 12, 13, 14, 15, 16, 17, 18, 19 20, 21, 22, 23, 25 28,29, 30, 31, 32, 33, 34, 35, 36, 37, 38 39, 40, 41, 42, 43, 44, 45, 46, 47, 48

Por ejemplo: Necesito analizar un cuadro que tiene las variables de edad y sexo.

- Para la variable edad determino primero que corresponde a la escala cuantitativa, por lo tanto, puede elegir entre los indicadores 6, 8, 9, 11, 12, 13, 14, 15, etc…, que significan razón, porcentaje, prorrateo, media, mediana, etc…, respectivamente. En el caso concreto si me interesa analizar el porcentaje con que se distribuyen las edades; por lo tanto, deberé contemplar una columna más en el cuadro para presentar los mismos; por otro lado, el análisis más importante será el sacar promedio de edad de la distribución el cual se obtendrá por separado y para utilizarse, dentro de la redacción, del análisis que se haga al cuadro.

Como se obtiene media, como medida de tendencia central, es lógico que se tendrá que obtener las medidas de dispersión para el adecuado análisis y por supuesto el coeficiente de variación (14) para determinar si fue adecuado o no el uso de la media o debiera usarse la mediana. De esto último se estudiara más adelante, ahora lo importante será saber manejar la guía de selección de indicador o analizador.

b) para la variable, sexo, se determina primero que corresponde a la escala CUALITATIVA: de todas las diferentes alternativas que tengo para analizar esta variable, según la guía, solo me interesa el porcentaje (8)

Análisis

VALORES RELATIVOS

DATOS DE ESCALA CUALITATIVA

NOTA: es probable que en algunos ejemplos, de los que se manejan en este documento sean absolutamente ficticios; lo importante es que tengamos algunos valores para ejemplificar.

Razón

Es la relación que existe entre 2 cantidades

Al realizar la división entre dos valores, se obtendrá un cociente, con esto se hace más fácil la comprensión de este valor en relación al otro, para una interpretación más clara

Formula: $\frac{a}{b}$

Ejemplo:

Supongamos que en un grupo se encontraron 68 personas de las cuales 42 eran mujeres y 26 hombres

42/26=1.62

Esto significa que por cada hombre presente, en este grupo, había 1.62 mujeres. Esta es la relación de un sexo en relación al otro.

Proporción

Es la igualdad o relación que existe entre dos razones.

Nos sirve para apreciar si es poco, o basta señalar la diferencia que existe entre dos cocientes de las razones que se comparan y con esto se aprecia la proporción que existe entre ambas.

Formula: $\frac{a}{b} = \frac{c}{d}$ o a:b::c:d

Ejemplo:

En el mismo grupo se encontraban individuos de diferentes edades pero destacaremos que, de 27 años, había 21 personas, y de 24 años se encontraban 13 personas.

Ahora comparo la proporción de las 2 variables (la edad y sexo) y aplico la fórmula: para esto debo tomar las 2 "razones" de ambas variables: 42 y 26, para sexo; 21 y 13 para la edad.

21:13 :: 42 : 26 o $\frac{21}{13} = \frac{42}{26}$

Podemos interpretar que: “en ese grupo, se presentó el caso de que la misma proporción de hombres y mujeres (1.62), fue para los individuos de 27 y 24 años (1.62)”.

8.-Porcentaje %

Es la parte proporcional, en relación a 100, que representa una cantidad, en relación a otra.

Es la expresión que facilita la interpretación: en virtud de que, comparar el valor absoluto con otro, sobre todo si es de cierta magnitud, no se4ria apreciado tanto, como cuando relacionamos una cifra comparándola con 100. Esto nos facilita la apreciación de un valor con respecto a otro, expresado en proporción de 100. Si nos damos cuenta esto sería, en su conjunto, una proporción ya que una razón seria el conjunto de 100 y la parte proporcional de estos.

Formula:

Es la aplicación de la regla de 3 simple o la misma que acabamos de ver en la proporción, nada más que ahora una de las dos “razones” se refiere exclusivamente a 100.

Para esto es recomendable que en uno de los miembros de la igualdad se coloquen los valores porcentuales y en el otro se coloquen los valores absolutos o cantidades.

Cantidad	%
Total	100
Parte	X

O T : p :: 100 : X

Notamos que los extremos están dados por: T y X, y los medios son p y 100. Por otro lado recordamos que: “el producto de los medios es igual al producto de los extremos”; por lo tanto estamos en posibilidades de encontrar cualquier valor de esta ecuación; así mismo, cualquier porcentaje que necesitemos. Basta multiplicar los 2 medios que conocemos (p)(100) y el producto lo dividimos entre el extremo que conocemos (T) -----(p)(100)/T.

En otras palabras: “el total (de una cantidad) es al 100 (total del porcentaje); como una parte (de la cantidad) es a una parte de 100”.

Ejemplo:

El censo poblacional del Estado de México, para 1980, fue de 7, 564,335 habitantes los cuales están distribuidos, por grupos de edad, como se relaciona en el siguiente cuadro.

NOTA: deseo aclarar que la tabulación (cuadro) estará presentada con la columna de % y que la forma de obtenerse se explica a continuación del cuadro.

Edad	Numero	%
‹1	257 189	3.4
1-4	892 591	11.8
5-14	2 276 865	30.1
15-44	3 252 664	43.0
45-64	642 968	8.5

65 y mas	242 058	3.4
Total	7 564 335	100.0

Para obtener el primer%

$$\frac{7564335}{257189} = \frac{100}{x}$$

(257 1899)(100)/7 564 335)=

=3.40002057677=

3.4 También podría ser:

(257 1899 / 7 564 335)100= 3.40002057677 = 3.4

La misma operación se deberá hacer para obtener cada uno de los porcentajes; sin embargo, se puede obtener el mismo resultado abreviando la mecanización y evitándose tantas operaciones, utilizando PORRATEO.

9.- Prorrateo

Proceso mediante el cual se distribuye una cantidad total (nueva), en la misma proporción con que se presenta otra cantidad, pero que esta última se encuentra desglosada.

Nos sirve para poder desglosar una nueva cantidad total, obtenida mediante algún mecanismo y que se ignora su comportamiento o desglose, requerimos tomar a otra que ya se encuentra desglosada y que nos servirá de base para distribuir la nueva cantidad total en la misma proporción, con que se presenta este cuadro base, desglosado.

Ejemplo:

Supóngase que después de proyectar la población para el año de 1986, al 30 de junio, se calculó que habría 11,272,482 habitantes; queda ahora el problema de su distribución etarea y por sexos. Como no tenemos otra base para calcularla, se obtiene basándose en la misma distribución que presentaba en el momento del censo, suponiendo que no haya cambios significativos en una década; por otro lado no se obtiene conocimientomás real del comportamiento en su distribución hasta el siguiente censo poblacional, por lo tanto se hará la distribución con la única base que tenemos, del censo de 1980.

Se sugieren los siguientes pasos:

1° se obtiene una constante

$$\frac{cantidad\ a\ la\ que\ quiero\ \text{llegar}}{cantidad\ que\ tengo\ como\ base} = k$$

$$\frac{11272482}{7564335} = 1.4902145$$

2°se multiplica la constante (k), por cada cantidad parcial

(1.4902145)(257189)= 383267

(1.4902145)(892591)= 1330152 etc….

Se continúa con todas las demás, para hacer el cuadro

1980

Edad	Numero	%
‹1	257 189	3.4
1-4	892 591	11.8
5-14	2 276 865	30.1
15-44	3 252 664	43.0
45-64	642 968	8.5
65 y mas	242 058	3.4
Total	7 564 335	100.0

1986

Numero	%
383 267	3.4
1 330 152	11.8
3 393 018	30.1
4 847 167	43.0
958 160	8.5
360 718	3.4
11 272 482	100.0

TASA

Relación que existe entre un determinado evento y la población en que acontece; en un lugar y tiempo determinado, y referido a una constante (10, 100, 1000….. etc.)

Sirve para determinar el riesgo de que suceda un evento, sobre una población; y así, como comparamos una cantidad, en relación a 100; es más comprensible decir que hubo 17 defunciones por cada 1000 habitantes, que mencionar que hubo 201 064 defunciones en 11 827 340 habitantes.

Formula:

$$\text{Tasa} = \frac{\text{fenómeno (número de veces)} \times 10}{\text{Población en la que ocurre *}}$$

*población expuesta al riesgo, a mitad de periodo

NOTA: siempre se tomara en el denominador, solo a la población expuesta al riesgo o ala propia del fenómeno, y propia del lugar.

Las tasas pueden ser:

1. General o cruda: todas las causas o toda la población
2. Especifica: determinada causa y/o parte de la población

Principales tasas:

- Mortalidad
- Morbilidad
- Natalidad
- Densidad de población
- Letalidad
- Incidencia
- Prevalencia

Análisis de datos para escala cuantitativa

Ẋ.μ MEDIA

Es el valor que tendrían todas las observaciones si no hubiera diferencia entre ellas.

Sirve para analizar la información en donde se pueda apreciar con una cifra y tener una idea más o menos, aproximada; mediante esta cifra, de cómo es el grupo; ya que al observar los valores de todos, se resalta o aprecia. En cambio, tomando una sola cifra se tiene rápidamente una idea general.

La media (promedio aritmético), la mediana y la moda, son medidas de tendencia central, que describen a la serie a través de un valor; su nombre genérico, será el del promedio.

Existen, así mismo, medidas de dispersión; tales como: varianza, desviación estándar y error estándar, para la media, y cuartil, decil y percentil, para la mediana.

Existen características que determinan cuando se debe utilizar la media, mediana y moda. Esto sucede, por ejemplo, cuando dentro de la serie existen valores extremos que pueden afectar a la media; esto se identifica aplicando el COEFICIENTE DE VARIACION (C.V) y si este es mayor de 30% entonces no se emplea la media; si no la mediana.

A) FORMULA (serie simple):

$$\dot{X} = \frac{£Xi}{n}$$

En donde:

£= sumatoria

Xi= variable

n= número de observaciones o elementos observados

Ejemplo:

En una encuesta a 5 alumnos, en que se preguntaba el número de pacientes, al mes; se encontró que informaban:

n alumno	Xi Pacientes
A	6
B	2
C	12
D	5
E	7
SUMA	32

$$\dot{X} = \frac{32}{5} = 6.4$$

Lo que significa que las 5 alumnos tienen un promedio (media) de 6.4 pacientes al mes, o sea, 6 al mes.

Existe otra fórmula y desarrollo para obtener la media, cuando se trata de datos en una serie que se agrupa

Supongamos, con el ejemplo anterior, que ahora las alumnos estudiadas fueron 105 y que ahora la distribución la hago de tal manera que enlista, ya no individualmente; si no que hago la relación, escribiendo el valor (Xi9 y la frecuencia (fi):

Xi Núm. De pacientes	Fi Núm. de alumnos	(Xi)(Fi)
1	5	5
2	8	16
3	11	33
4	13	52
5	17	85
6	21	126
7	12	84
8	9	72
9	5	45
10	3	30
11	1	11
Suma	105	559

B) FORMULA (serie agrupada)

$$\dot{X} = \frac{£Xifi}{£fi}$$

En donde:
£= sumatoria
Xi= valor observado
Fi= frecuencia con que se presentó el valor observado
£fi= número de elementos estudiados
XiFi= producto de multiplicar el valor por su frecuencia

Así pues:

$$\dot{X} = \frac{559}{105} = 5.32$$

Concluimos que el promedio de pacientes, al mes, de las 105 parejas estudiadas es de 5.32, o sea, 5.

Evidentemente este sería el número de pacientes que tendrían todas las parejas, si no hubiera diferencia entre ellas; sin embargo, la realidad es que no todos se comportan de la misma manera y por lo tanto se pueden hacer análisis, al respecto para poder describir con más precisión al grupo. Así pues, se nos ocurre pensar en que si existen diferencias; si este grupo es parecido (homogéneo). Para esto nos servirá utilizar las respectivas medidas de dispersión de cada promedio, tales como desviación estándar, cuartil, decil, percentil, etc….

Observamos que hemos significado a la media como $\bar{x}$; también, se puede escribir corno µ (mi u). La diferencia entre una y otra estará en que cuando se tienen los datos dirceamente. El resultado se escribe como X; en cambio cuando los dalos se refieren a la población completa, y que no se manejaron directamente todos; o quizá se está infiriendo (calculando) sobre toda la población o universo, el resultado se significa con: µ

$\bar{x}$ - µ

Antes de concluir con la media, habrá que apuntar. Que si la serie fuera agrupada, por clases, entonces no tendríamos un valor observado; si no una serie de valores agrupados en una clase. En este caso deberemos obtener un s6lo valor (Xi), y éste será Precisamente, EL PUNTO MEDIO DE CLASE.

EJEMPLO: supongamos la clase 35 - 39 años, en este caso el valor único estará dado por: 35 + 39/2 tt 37; 37 será (Xi).

Se habrá notado que en los cuadros presentados se han incluido Xi, f'i, (Xi fi), etc. Obviamente estos símbolos los usaremos para desarrollar los cálculos del análisis. Es claro que para la presentación del trabajo, sólo se harán los encabezados de las columnas necesarias, todas las demás columnas se habrán hecho para el desarrollo de los cálculos.

Ejemplo: Con los datos anteriores, las columnas a presentar Serian solamente:

Núm. De pacientes	Núm. De parejas

Md MEDIANA

Es el valor que divide a la serie en dos parles iguales.
Se puede considerar como valor promedio

En esta Medida, de tendencia central, no se toman en cuenta a todas las observaciones, para obtener el valor promedio, como sucede con la media. Sólo se toma el valor de la, o las, observaciones que se encuentran a la mitad de la serie, una vez que se han ordenado progresivamente en forma ascendente o descendente y éste será el valor promedio.

A.- FORMULA (Serie simple o compuesta gradual):

1o. Se ordenan los valores de las observaciones, en forma ascendente o descendente.

2o. Se busca el valor que se encuentra a la mitad de la serie, para esto se aplica:

$$MD. = \frac{N+1}{2}$$

Si a la mitad de la serie, coincidiera que no se encuentra valor único, por ser decimal el lugar que resulto de n + 1/2; entonces se tomarán los 2 valores centrales y se promedian.

3o. Una vez que se ha encontrado el lugar central; se observa cuál es el valor de la observación que se encuentra en ese lugar y éste será el valor promedio (valor medial).

B. - EJEMPLO:

Con el ejemplo, antes citado, de la frecuencia coital, n = 5 parejas

Pareja	**Pacientes**
A	6
B	2
C	12
D	5
E	7

1º. 6, 2, 12, 5, 7; se ordenan: 2, 5 , <u>6</u>, 7, 12

2º n + ½ ; 5 + ½ = 3

3 es el lugar donde se encuentra el valor medial

3º en ese caso la observación numero 3 corresponde al valor 6; lo que significa que el promedio mensual de pacientes, de las 5 parejas, es de 6.

B.- Formula (serie de clases y frecuencias):

$$MD. = 1 + \left[\frac{\dfrac{\sum^{N} fi}{2} - \sum^{m-1} fi}{Fm} \right]$$

En donde:

1 = Límite inferior de la clase medial, con límite verdadero (la clase en donde se encuentra el valor de la observación que está la mitad de la serie).

Para encontrar la clase: $\dfrac{\sum fi}{2}$

$\dfrac{\sum^{n} fi}{2}$ = Suma de las frecuencias de toda la serie entre 2.

$\sum^{m-1} fi$ = Suma de las frecuencias, acumuladas, hasta la clase anterior a la que contiene a la Md.

fm = Frecuencia simple de la clase que contiene la Md.

I = intevalo de la clase.

Ejemplos:

Xi Pacientes	Fi Parejas	fi Acumulada
1	5	5
2	8	13
3	11	24
4	13	34
5	17	54
6	21	75
7	12	87
8	9	96
9	5	101
10	3	104
11	1	105
Suma	105	-

Md. = 105/2 = 52.5: el lugar 52.5, se encuentra en el grupo que acumula desde los 37 hasta los 54, el lugar en donde se encuentra el valor 5. Las 105 parejas tienen un promedio de 5 pacientes mensuales.

En el ejemplo anterior, no fue necesario aplicar toda la fórmula completa, por ser puesta gradual (un solo valor; y no una clase), bastó buscar el lugar e inmediatamente se obtuvo el valor promedio.

Ahora veamos un ejemplo en donde se manejen clases: En 84 trabajadores, se desea obtener el promedio de jornal diario.

1=19.5

Clase Jornal $	Fi Número	Fi Acumulada
01 – 9	3	3
10 – 19	13	16
20 – 29	32	48
30 – 39	16	64
40 – 49	10	74
50 – 59	2	76
60 – 69	8	84
Suma	84	-

m-1
Σf1

Clase media Σ f1/2 84/2 = 42 el lugar 42 se encuentra en la f1. Acumulada de 17 hasta 48

fm

$$MD = 19.5 + \frac{84/2-16}{32}\ 10 = 19.5 + \frac{42-16}{32}\ 10 \quad =$$

$$=19.5\ \frac{26}{32}\ 10 = 19.5 + \frac{260}{32} \quad =$$

=19.5 + 8.125 = 27.62

El grupo de 84 jornaleros gana $ 27.62 en promedio.

Comparemos este mismo ejemplo obteniéndolo con la media

CLASE	**FI**	**C.C.**	**XIFI**
01 - 09	3	5.0	15
10 – 19	13	14.5	188.5
20 - 29	32	24.5	784
30 – 39	16	34.5	552
40 – 49	10	44.5	445
50 – 59	2	54.5	105
60 – 69	8	64.5	516
SUMA	**84**	**-**	**2605.5**

$\bar{x}$ =2 605 .5 / 84 = 31.017857 = 31.02

Notamos que la media es igual a 31.02, la mediana es igual 27.62

Mo, MODA

Es el valor de máxima frecuencia en la serie

Pocas veces se utiliza; sin embargo, cuando un valor se repite muchas veces y es evidente que ese valor podría describir al grupo, se le considera como valor promedio de la serie. No importa que sea de serle simple o compuesta, en ésta última, se tomaría el centro de clase (promedio de la clase); así sería el promedio obtenido con la moda.

EJEMPLO: En el de las 105 parejas, se tomará simplemente el valor que más se repite; seria el 6, que se repite 21 veces y es el de mayor frecuencia.

Hasta aquí hemos visto las 3 medidas de tendencia central, que sirven para analizar datos de escala cuantitativa, continúa o discreta; veamos ahora las medidas de dispersión.

Medidas De Dispersión (Variación)

Para el análisis de los datos de escala cuantitativa, no basta saber el valor central o promedio, para describir la distribución que tengamos enfrente, hace falta conocer varias cosas más y tener los suficientes argumentos para concluir con un adecuado análisis.

Las medidas de dispersión. Que también se llaman de variación, son valores que indican cómo se comporta la distribución de la serie, al observar si es homogénea, compacta, agrupada, muy parecidos; o dispersa, heterogénea, esparcidos, muy diferentes.

Sirven, en primer lugar, para decidir cuál Analizador (media o mediana) se tiene que aplicar, en segundo lugar, para encontrar valores limites, entre los cuales se pudieran dar las variaciones que se present.an en el grupo de estudio.

Consideremos algún caso:

Supongamos que las parejas que hemos estudiado durante mucho tiempo, han manifestado que la frecuencia de pacientes mensual, ha sido de 16; pacientes; este valor promedio, nos describe "a las parejas" en forma general; al meternos a buscar mayor información, nos encontramos con que ha sido muy variable en estas parejas, pues algunas refieren que a veces pasar, 3 meses sin tener algún paciente; por otro lado algunas, Manifiestan, Hasta 67, al mes. Esto nos habla de una gran "variación." entre las parejas; si en lugar de concretarnos a decir quo el promedio es de 16, dijéramos: " el promedio es de 16, pero la MAYORIA está entre 11 y 21 pacientes al mes ... ", estamos encontrando que existe una diferencia promedio de 5 más o menos de 16, el número de pacientes que tienen al mes.

Con la idea anterior nos damos cuenta que nos falta mayor información del comportamiento de los datos en nuestra distribución, para integrar adecuadamente nuestro análisis, que independientemente de ideas técnicas, le agregaremos los conocimientos e interpretaciones necesarias para hacer nuestras conclusiones.

Medidas de dispersión (variación)
a).- Rango
b).- Coeficientemente de variación (C. V.)

Para la media:
c).- Varianza (S^2, o^2)
d). - Desviación estándar (S, o)
e).- Error estándar (E.E., ox)

Par a la mediana;
f).- Cuartil(a) (Q)
g). – Decil(a) (D)
h).- Percentil(a) (P)

C. V. COEFICIENTE DE VARIACION

Es la medida que sirve para observar si existe mucha o poca dispersión entre los valores, con respecto a su media.

Sirve para que, mediante este estimador, se decida si es conveniente analizar los dalos a través de la media o preferir la mediana.

La decisión de una u otra medida de tendencia central se basa, en que pudiera haber valores extremos que afectan a la serie y no se describa adecuadamente a través de la media. En este caso, aun cuando no se consideran todos los valores, en su comportamiento dentro de la serie, puede ser descrita con la mediana.

Para encontrar si existe mucha dispersión se aplica el coeficiente de variación y si éste es mayor que 30 %, se debe usar la mediana. Si es poca dispersión (menos del 30 %). Se utilizara media.

FORMULA:

$$C.V. = \frac{S}{X} 100$$

Si el C. V. Es < 30%, hay poca dispersión y la X representa adecuadamente a la distribución. Usar MEDIA.

Si el C. V. Es >30 %, Existe mucha dispersión, y la X, NO representa adecuadamente a la dislribuci6n. Usar MEDIANA.

Con el comentario hecho al inicio de este capítulo, nos hemos dado cuenta de que se puede determinar un valor central para describir a la serie; sin embargo, también notamos: que cada individuo tiene su propio valor, dentro de la serie, que lo hace diferente del promedio general o puede también coincidir. Para medir esta dispersi6n, desviación o diferencia con respecto a su media; claro está que se puede ir observando uno por uno, pero esto resultaría muy problemático, a la par, que poco podríamos apreciar, sobre todo cuando son numerosas las observaciones..

Para poder apreciar estas diferencias tenemos que recurrir a obtener otro valor que indique su "Diferencia o desviación, en promedio" de todo el grupo.

a).- Si observamos el ejemplo de las 5 parejas, nos damos cuenta que el valor promedio es de 6.4 pacientes al mes. Así, la diferencia de cada pareja, con respecto a su media es de:

A = -0.4 = (6 - 6.4); B= -4.4 = (2 - 6.4); C = +5.6 = (12 - 6.4)

D = -1.4 = (5- 5.4); E = + 0.6 = (7 - 6.4). Hemos visto que respetivamente se diferencian, con respecto a la media en:

-0.4, -4.4, +5.6, -1.4 y +0.6. Esto es, (Xi - $\bar{X}$)= diferencia de cada valor con respecto a su media. Aun así no se puede apreciar adecuadamente y recurriremos a obtener algún valor único que nos ayude a describir esa diferencia, positiva o negativa.

b) Si sumáramos algebraicamente (positivos y negativos), las diferencias, es lógico que obtengamos un valor cercano al cero, o cero en la suma. Para evitar este resultado, Tendremos que elevar al cuadrado cada diferencia, con esto se pierde lo negativo y nos resulta la "Diferencia cuadrática" de cada una; ejemplo: (-0. 4) (-0. 4) = 0.16. Esto es (Xi-$\bar{X}$) e "diferencia de cada valor con respecto a su media, elevado al cuadrado."

Los valores quedarían, respectivamente: 0.16, 19.36, 31.38, 1.96 y 0.38. Esta es la diferencia cuadrática de cada una.

c) Sigue siendo un conjunto de diferencias (ahora cuadrática); es necesario que obtengamos un valor "único", promedio que nos diga la diferencia del grupo. para esto se suman estas diferencias cuadráticas Σ: (Xi-$\bar{X}$) ² = "sumatoria de la diferencia de cada valor, con respecto a su media, elevado al cuadrado" : 0.16 + 19. 36 + 31. 36 + 1. 96 + O. 36 = 53. 2. Este resultado se divide entre n (número de individuos o elementos estudiados), así tenemos que:

$$S^2 = \frac{\Sigma (xi - x)^2}{n} = \frac{53.2}{5} = 10.4$$

10.64 es la diferencia cuadrática de cada valor con respecto a su media llamada VARIANZA.

Hasta aquí la varianza "poco" nos puede decir para la idea que buscábamos; ésta se usará más bien para estudios más profundos., dentro de la estadística inferencial cuando se trabaje con “análisis de varianza”

d) Lo que buscamos es conocer "la diferencia, promedio, de cada valor con respecto a su media" y esto es: la desviación estándar Para obtenerla basta con sacar la raíz cuadrada de la varianza.

Así:

$$S= \sqrt{\frac{\Sigma (Xi - X)^2}{n}} = \sqrt{\frac{53.2}{5}} = 10.64$$

S = 3.261901286 = 3.3

Concluimos diciendo que la " desviación estándar (diferencia promedio de cada valor, con respecto a su media) es de 3.3 pacientes.

La idea general de este ejemplo quedaría así:

"En las 5 alumnos, estudiadas se encontró que el promedio mensual de pacientes es de 6.4 y que se diferencian, en promedio, en 3.3”

Surge ahora la pregunta, es poca o mucha la dispersión de una serie?

Ya habíamos anunciado como medida de dispersión al coeficiente de dispersión y si lo aplicamos encontramos:

C.V= $\frac{S}{\bar{X}}$ = 100 = $\frac{3.3}{6.4}$ = 100 = $\frac{300}{6.4}$ = 51.5625

S^2 , o^2

VARIANZA.

Es la diferencia (desviación) cuadrática promedio de los datos, en relación a su valor promedio (media)

FORMULAS:

SERIE SIMPLE

$S^2 = \frac{\sum (X1-X))2}{n}$ Ó $S^2 = \frac{1}{n}\ (\sum_{X1} 2 - n\ \ \bar{X}^2\ )$

SERIE AGRUPADA

$S^2 = \frac{\sum (X1-X)2\, fi}{\sum fI}$

En donde :

$\sum$ = Sumatoria

$(X1 - \bar{X})2$ = Diferencia de cada valor, respecto a su media , elevado al cuadrado

n: Número de individuos o elementos que conforman la serie o distribución

$\sum_{X1} 2$ = Sumatoria de las variables al cuadrado, se eleva cada variable al cuadrado y luego se suma.

$\bar{X}^2$ = El valor de la media elevado al cuadrado.

fi = Frecuencia de cada variable.

$\sum (X1 - X)2\, fi$ = Sumatoria de las diferencias de cada valor , con respecto a su media elevada al cuadrado, por su frecuencia. Primero se obtiene la diferencia luego se eleva la cuadrado, en seguida se multiplica por su frecuencia, esto para cada variable luego se hace la suma total de estos productos

DESVIACION ESTANDAR

Es una medida de dispersión o variabilidad de los datos respecto a su media. Como con la media encontramos diferencia en la forma de significar la desviación estándar *S* cuando manejo los datos directamente de mi universo muestra *o* cuando lo infiero a la población total y o_x cuando se maneja dispersión en muestreo o probabilidad.

S= $\sqrt{\frac{\sum (X1-X)2}{n}}$ *Ó* *S*= $\sqrt{\frac{1}{n}\ (\sum_{X1} 2-n\ \bar{X}^2)}$

S= $\sqrt{\frac{\sum (X1-X)2\,fi}{\sum fI}}$ *S*= $\sqrt{S^2}$

Los elementos de las fórmulas son los mismos que los de la varianza, tengamos en cuenta que la única deferencia que existe es la obtención de la *S*, S^2, *C.V*, o_x.

SERIE SIMPLE:

N ALUMNO	X1 PACIENTES	X1-$\bar{X}$	$(X1-X)2$	$X1^2$
A	6	-0.4	0.16	36
B	2	-4.4	19.36	4
C	12	+5.6	31.36	144
D	5	-1.4	1.96	25
E	7	0.6	0.26	49
SUMA	32	----	53.2	258

Elementos necesarios:

n= 5 $\sum_{X1} 2$ = 258

$\sum X1$ = 32
$X1^2$= 40.96

$X1$ = 6.4

$\sum(X1-\bar{X})$ = 53.2

APLICACIÓN:

a) .- $S^2 = \frac{\sum (X1-\bar{X})2}{n} = \frac{53.2}{5}$ = 10.6

b) .- $S^2 = \frac{1}{n}\ (\sum_{X1} 2 - n\ \ \bar{X}^2\ ) = \frac{1}{5}\,[258 - 5(6.49)^2] = \ \frac{1}{5}\,[258 - 5(40.96)]$= 10.6

Observamos que con los dos procedimientos se obtiene el mismo resultado

Como la fórmula es la misma, solo que sacando raíz cuadrada, nos resulta:

S= $\sqrt{S^2}$ = $\sqrt{10.6}$ = 3.261901286 = 3.3

COEFICIENTE DE VARIACIÓN:

C.V= $\frac{S}{\bar{X}}$ = 100 = $\frac{3.3}{6.4}$ = 100 = 51.6%

ERROR ESTANDAR:

$o_X = \frac{S}{\sqrt{n}} = \frac{3.3}{\sqrt{5}} = \frac{3.3}{2.236068}$ = 1.4768 = 1.5

SERIE COMPUESTA

CLASE	fi	c.c	X1fi	X1-$\bar{X}$	$(X1 - \bar{X})2$	$(X1 - X)2\ fi$
01-9	3	5	15	-26.018	676.93	2030.69
10-19	13	14.5	188.5	-16.518	272.84	3546.92
20-29	32	24.5	784	-6.518	42.48	1359.38
30-39	16	34.5	552	3.482	12.13	194.08
40-49	10	44.5	445	13.482	181.77	1817.70
50-59	2	54.5	105	23.482	551.41	1012.82
60-69	6	64.5	516	33.482	1121.05	8968.40
SUMA	84	---	2605.5	----	----	19020.07

Elementos necesarios:

∑ fi= 84

$\sum(X1 - \bar{X})2$ = 19020.07

VARIANZA:

S= $\sqrt{\frac{\sum(X1-X)2\ fi}{\sum fI}}$ = $\frac{19020.07}{84}$ = 226.429404761 = 226.43

DESVIACION ESTANDAR:

$S = \sqrt{S^2} = \sqrt{226.43} =$ 15.047571 = 15.05

COEFICIENTE DE VARIACION:

C.V= $\frac{S}{\bar{X}}$ = 100 = $\frac{15.05}{31.02}$ = 100 = 48.5%

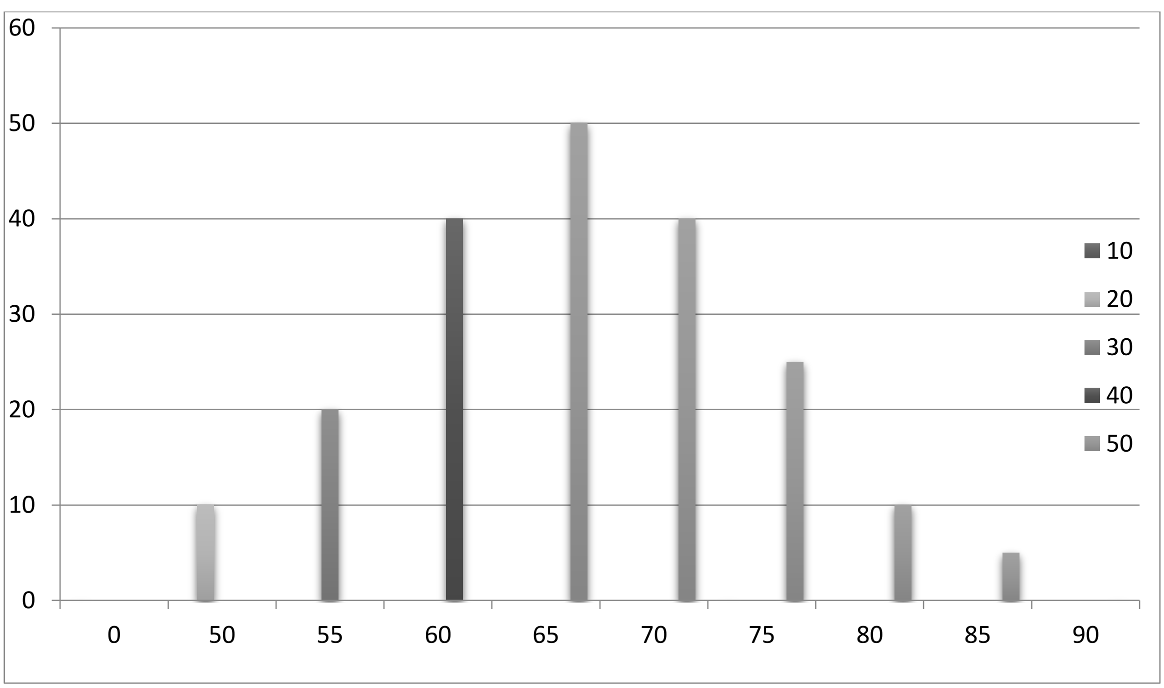
ERROR ESTANDAR:

$o_X = \frac{S}{\sqrt{n}} = \frac{226.43}{\sqrt{48}} = \frac{15.040757}{9.165151.}$ = 1.6417685 = 1.64

CURVA NORMAL DE GAUSSE

Aun cuando la curva normal se utiliza en la probabilidad, como la distribución normal , es interesante verla en este momento para completar la idea de la desviación estándar , se le conoce también como la campana de Gauss, porque cuando se tiene un gráfico de barras, bastones o histograma al hacer una cresta o seguimiento de los perfiles superiores de las barras se encuentra con la aproximación a una campana.

Imaginemos por ejemplo que tenemos diferentes medidas antropométricas de un grupo de individuos, al hacer el cuadro y especialmente la gráfica podemos percatamos que en ambos extremos de las ´´X¨ abscisa o variable independiente se encuentran los valores más bajos y a la derecha los valores más altos así encontramos que en cualquiera de estas descripciones o distribución encontramos pocas personas en las medidas más altas y los más se encuentran en el estándar o normalidad, esto es en el centro de la curva.



Si observamos la gráfica e imaginamos una campana que forma la propia grafica podemos entender que esta campaña forma en la superficie total debe contener la unidad, de tal manera que toda el área vale 1. Si dividimos está en partes, esto es en 3 + o hacia la derecha (+), o a la izquierda (-) ; y levantamos estas líneas hasta el arco de la curva, tendremos 6 áreas.

Cada una de estas líneas son llamadas unidades tipificadas, o precisamente desviaciones estándar No necesariamente en unidades enteras, evidentemente puede, y existen fracciones, puede haber; 1.96 desviación ó 2.68 etc.

Partiendo del centro (donde se encuentra la $\bar{X}$), abarcando entre 1 desviación estándar (S) + a la derecha o (-) izquierda, esperamos encontrar, dentro de esa área comprendida entre dos desviaciones estándar + - , se espera encontrar al 95.45% y entre 3S + - , se encontrara al 99.73% . Esta es la base de lo que se estudiara más adelante en la probabilidad.

Como ya se ha comentado, las medidas de dispersión para la mediana, son el cuadril Q , decil D y percentil P significa que una vez que hemos obtenido el valor promedio, de tendencia central, la mediana MD y para ayudarnos al análisis de la distribución es necesario buscar algunos valores entre los cuales s e encuentran la mediana y apreciar con más claridad si el grupo es disperso (heterogéneo) o compacto (homogéneo)

Q CUARTIL (A)

Valor que sirve para delimitar o indicar la diferencia que existe de los valores observados, con respecto a su valor medial promedio, considerándose en 4 partes.

La idea es similar a la de la desviación estándar;. Difiere en que aquí no se consideran para su determinación el comportamiento general de todos los valores de la seria, sino que solo algunos que sirven para marcar los límites entre los cuales se puede encontrar al 50% de todos los valores observados en la serie. Estamos hablando de un valor promedio (la mediana) y la tercera parte Q3 abarcando, con esto, un 50% central, y no un 60.97% como en la media.

FORMULAS:

SERIE SIMPLE:

$O_x = \frac{n+1}{4}$ k el resultado en el lugar

SERIE COMPUESTA O DE CLASES Y FRECUENCIAS:

$$O_X = 1_k + \left\{ \frac{\frac{\sum fi}{4} k - k - 4_{\sum fi}}{fm_k} \right\} 1_k$$

NOTA: K= Para no estar repitiendo el número de cuartil se desea buscar, se ha significado con K ; el cual va ir variando según se necesite; 1, 2 o 3.

En la fórmula:

1. = Límite inferior de la clase cuartilar con límites verdadero (la clase en donde se encuentra el valor de la observación que está en la cuarta parte de la serie).

Para encontrar la clase: $\frac{\Sigma \text{fi}}{4} \kappa$

$\frac{\overset{n}{\Sigma} \text{fi}}{4} \kappa$ = Suma de las frecuencias, de toda la serie entre 4 multiplicado por el cuartil respectivo.

$\overset{\kappa-1}{\Sigma} \text{fi}$ = Suma de las frecuencias, acumuladas, hasta la clase anterior a la que contiene al cuartil "k".

fm_k = Frecuencia simple de la clase que contiene al cuartil k.

i_k =Intervalo de la clase cuartilar.

D DECIL (A)

Valor que sirve para delimitar o indicar la diferencia que existe de los valores observados, con respecto a su valor medial (promedio), considerándose en 10 partes.

Es el mismo concepto del cuartil, sólo que ahora estamos dividiendo a la serie en 10 partes y no en 4 como fue en los cuartiles. Con esto podemos abrir o cerrar nuestros límites de "confianza" tanto cuando queramos; así pues, podríamos trabajar entre el decil 2 y el decil 8, lo que significará que estamos dejando a los extremos al 40%; o abarcando a un 60% de los valores observados entre D_2 y el D_8.

Formula: serán las mismas anteriores; se diferencia en que en lugar de Q, es "D" (decil) y la "k" se considera cada decima parte y no la cuarta parte.

P PERCENTIL (A)

Valor que sirve para delimitar o indicar la diferencia que existe de los valores observados, con respecto a su valor medial (promedio), considerándose en 100 partes.

Continuamos con el mismo concepto del cuartil y el decil; nada más que ahora estamos dividiendo a la serie en 100 partes o lugares, de tal manera que ahora podemos ubicar los límites que deseemos y por tanto, el porcentaje que queramos, según sea necesario; si, por ejemplo, deseo tener un 90% de “confianza”, deberá trabajar los datos entre el P_5y el P_{D5}; estos valores limitaran el 90% que deseo observar.

Nos habremos dado cuenta que las medidas de tendencia central (media y mediana), tienen algo en común, que ambas cuentan con medidas de dispersión que lo hemos entendido como valores intermedios, en toda la serie, que pueden servir de límites para poder describir el comportamiento de los valores, con respecto a su valor promedio. Si son más cerrados o abiertos estos límites, para detectar si existe mucha o poca dispersión; o si al grupo lo podemos describir como homogéneo o heterogéneo, respectivamente.

Fórmula: La misma que la del cuartil; se substituye, ahora la Q, por “P”; y la k, significa ahora, la centésima parte de la serie.

Ejemplo:

Con los mismos datos, con los que obtuvimos la mediana, haremos el ejercicio para obtener el cuartil (Q), el decil (D) y el percentil (P).

Clase Jornal $	f i Número	f i acumulada
1-9	3	3
10-19	13	16
20-29	32	48
30-39	16	64
40-49	10	74
50-59	2	76
60-69	8	84
Suma	84	--

Q_1, cuartil 1:

$$Q_1 = 19.5 + \frac{\left[\frac{84}{4} - 16\right] 10}{32} = 19.5\,\frac{21-16}{32}\,10 = 19.5 + \frac{50}{32} = 19.5 + 1.5625 = 21.0625 = 21.06$$

Q_3, cuartil 3:

$Q_3 = 29.5 + \frac{\frac{84}{4}3 - 48}{16} 10 = 29.5 + \frac{(21)3-48}{16} 10 = 29.5 + \frac{63-48}{16} = 29.5 + \frac{15}{16} 10 = 29.5 + \frac{150}{16} = 38.875$

D_2 decil 2:

$D_2 = 19.5 + \frac{\frac{84}{10}\ 2 - 16}{32} 10 = 19.5 + \frac{(8.4)2-16}{32} 10 = 19.5 + \frac{16.8-16}{32} 10 = 19.5 + \frac{0.8}{32} 10 = 19.5 + \frac{80}{32}$

$= 19.5 + 2.5 = 22$

D_8 decil 8:

$D_8 = 39.5 + \frac{\frac{84}{10}\ 8 - 64}{10} 10 = 39.5 + \frac{(8.4)8-64}{10} 10 = 39.5 + \frac{67.2-64}{10} 10 = 39.5 + \frac{3.2}{10} 10 = 39.5 + \frac{32}{10}$

$= 39.5 + 3.2 = 42.7$

P_{10} Percentil 10:

$P_{10} = 9.5 + \frac{\frac{84}{100}\ 10 - 3}{13} 10 = 9.5 + \frac{(0.84\ 10-3}{13} 10 = 9.5 + \frac{8.4-3}{13} 10 = 9.5 + \frac{5.4}{13} 10 = 9.5 + \frac{54}{13}$

$= 9.5 + 4.15 = 13.65$

P_{90} Percentil 90:

$P_{90} = 49.5 + \frac{\frac{84}{100}\ 90 - 74}{2} 10 = 49.5 + \frac{(0.84)\ 90-74}{2} 10 = 49.5 + \frac{75.6-74}{2} 10 = 49.5 + \frac{1.6}{2} 10 = 49.5 + \frac{16}{2}$

$= 49.5 + 8 = 57.5$

INTERPRETACION:

Así como podíamos decir, en relación a la media; que esta se puede encontrar entre tal y cual valor, tomando en cuenta como límites, a la desviación estándar; de la misma manera podremos señalar límites, entre los cuales se encuentran la mediana, sólo que ahora con el Cuartil (Q), Decil (D), o Percentil (P); que, como sabemos, son las medidas de dispersión de la mediana.

Recordemos que en este ejemplo, la Mediana es de $27.62.

Decimos que los jornaleros ganan, en promedio 27.62, y que deberían estar, el 60% de ellos, entre 20.11 y 42.70.

Mencionamos el 60%; aproximadamente, porque al estar entre el decil 2 y el decil 8, quedan hacia afuera (cola izquierda y cola derecha), 20 y 20% (sumando el 40%); por lo tanto, consideramos, hacia el centro al 60%.

Entre 21.06 y 38.87, se debiera encontrar al 50% de los jornaleros; puesto que estas cantidades están comprendidas con los cuartiles 1 y 3 (Q_1 y Q_3), 25 y 25% quedan afuera.

Entre 13.65 y 57.50, es de esperarse que se encuentren el 80%, de los jornaleros, puesto que esas cantidades están limitando el percentil (P) 10 y 90, respectivamente, luego entonces, quedan en los extremos 10 y 10 (20%).

M MOMENTO

Es una serie de pasos que permiten obtener medidas sumarias de conjuntos de datos.

Partiendo de que la $\bar{X}$ es el primer momento, en una sucesivamente. Sirve de base para calcular el sesgo y la curtosis.

21.- SESGO

Es el grado de asimetría de una distribución.

Se califica por la dirección que toma la cola de la curva, al graficar. Puede ser sesgo positivo, si la cola se agranda (inclina) hacia la derecha; sesgo negativo, si la cola se inclina o agranda a la izquierda. Si no existiera sesgo o inclinación, hacia algún lado, se habla de una curva simétrica.

CURTOSIS

Es el grado de apuntamiento o altura de la curva, en una distribución.

Estas tres últimas medidas, si nos damos cuenta, nos sirven para poder describir a la curva que resulta de haber graficado los datos; puesto que si hacemos la gráfica de barras o histograma, según el caso, y fuéramos marcando por la parte superior de las alturas de las columnas del gráfico, nos va a resultar algo que semeja, en mayor o menor grado, a una campana.

Supongamos que describimos las estaturas de un grupo de individuos adultos, de cualquier grupo; al graficar los datos y marcáramos por la parte superior de las columnas del histograma, encontraríamos esa figura de campana. Observamos pocos individuos en el extremo izquierdo, de poca estatura, y pocos, de mucha estatura, en el extremo derecho, y subiendo hacia el centro, por ambos lados, esta curva, hasta encontrarnos que la mayoría (cerca del 70% se encuentran precisamente en la parte central).

Si la muestra fuera pequeña, tal vez no nos diera esta “campana”; pero mientras más grande sea la muestra, la figura de campana va apareciendo con mas claridad, conforme se va cerrando, hacia el centro; por consecuencia, irá subiendo por la parte central.

Curva alta y colas anchas: Leptocúrtica; curva baja y colas finas: Platicúrtica.

Esta descripción de los momentos, sesgo y curtosis, son los que se denominan “Pruebas de Normalidad”, o sea, que tanto se va acercando la descripción del grupo, como para encajar en una CURVA NORMAL.

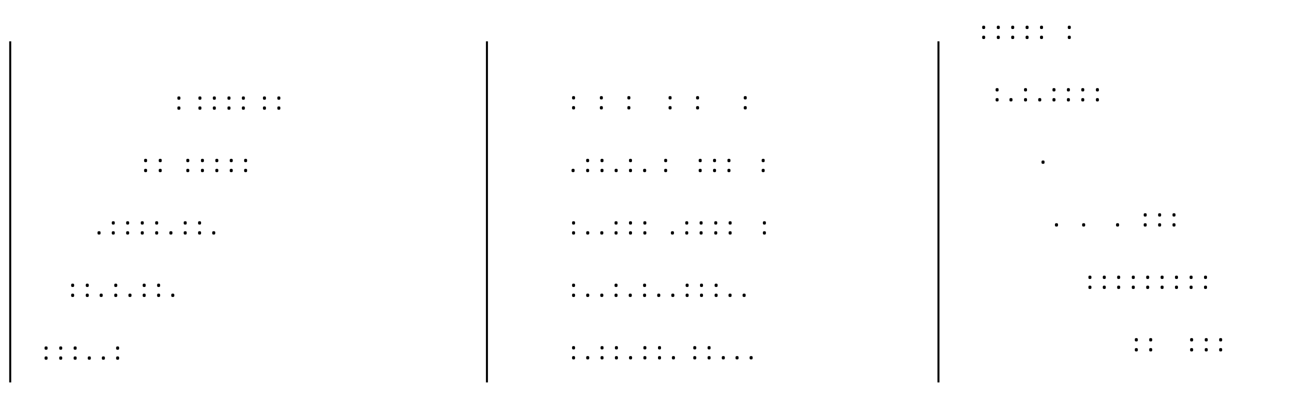
De aquí que en algún momento podemos decir que si la serie, es simétrica, “normal”, la media, mediana y moda, deben coincidir; si no es así, podrían quedar en diferente ubicación y siempre, en este caso, encontraríamos a la moda en la parte mas alta de la curva; la mediana en la parte central, y en uno de los extremos a la media, dependiendo del sesgo que tenga la curva.

|° INDICE DE CORRELACION

Es la medida que sirve para determinar que tanto se correlaciona o depende una variable, respecto a la otra.

Sirve para observar el grado de relación que existe entre las variables dependiente (Y) e independiente (X). Antes de tirar la línea de regresión es conveniente determinar si existe correlación, entre las variables, mediante el coeficiente de correlación.

En algunos momentos podemos encontrar datos que están dependiendo del comportamiento de otros. Podríamos mencionar, como ejemplo, que a mayor edad dene esperarse mayor estatura; adulta, ya no se da, podemos resumir que la estatura está relacionada con la edad, en el primer caso; en el segundo caso, no existe correlación. Al graficar se puede encontrar la siguiente presentación:



+ O ----

FORMULA:

$$|° = \frac{\frac{\Sigma X'Y'}{n}}{x^{\partial} \quad y^{\partial}}$$

En donde:

Σ = Sumatoria

n = número de observación

X´= X -$\bar{X}$

Y´= Y - $\bar{Y}$

$\bar{X} = \frac{\Sigma X}{n}$

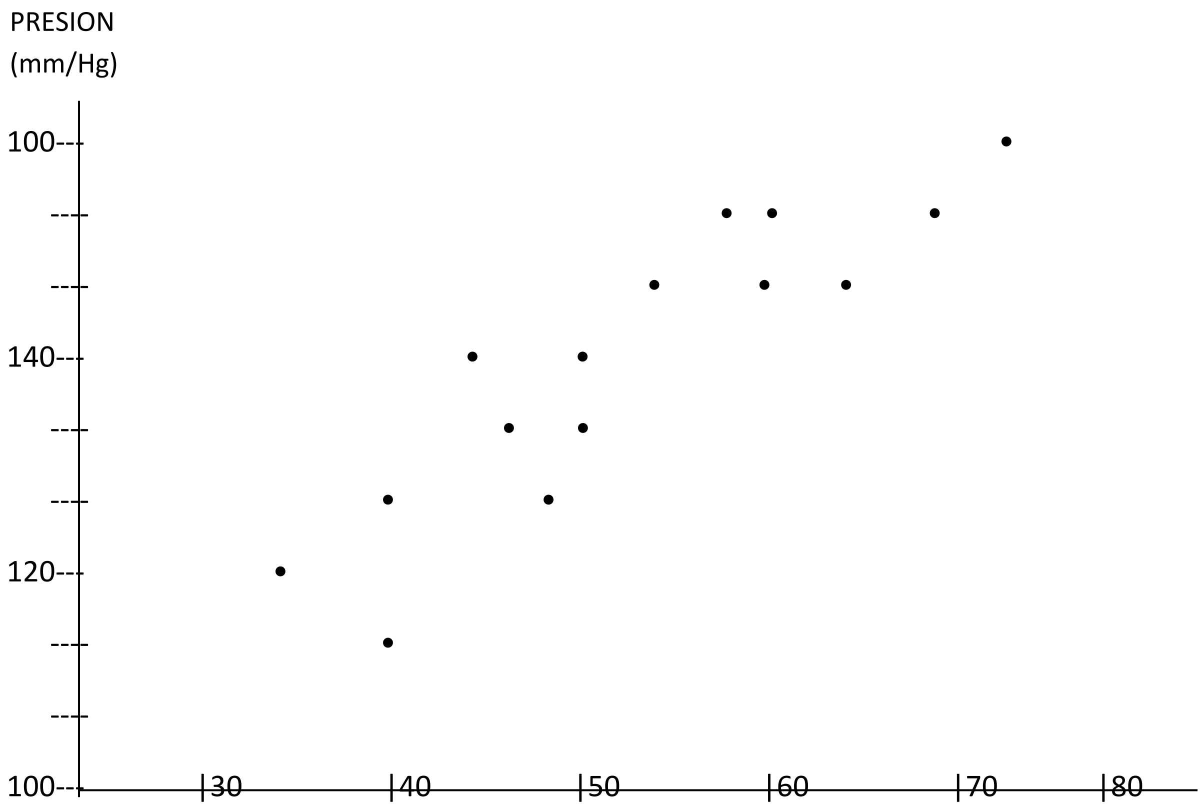
$\bar{Y} = \frac{\Sigma y}{n}$

$\partial x = \sqrt{\frac{\Sigma (x - \bar{X})^2}{n}}$

$\partial y = \sqrt{\frac{\Sigma (y - \bar{Y})^2}{n}}$

EJEMPLO:

Edad	Presión mm/Hg	Sx		Sy		X´ Y´
X	Y	X -$\bar{X}$	(x – $\overline{X})^2$	Y - $\bar{Y}$	(Y - $\bar{Y})^2$	(X –$\overline{X}$) (Y - $\bar{Y}$)
36	118	-16.33	266.67	-22.33	498.63	364.65
38	115	-14.33	205.35	-25.33	641.61	362.98
42	125	-10.33	106.71	-15.33	235.01	158.38
42	140	-10.33	106.71	-0.33	0.11	3.41
47	128	-5.33	28.41	-12.33	152.03	65.72
49	145	-3.33	11.09	4.67	21.81	-15.55
55	150	2.67	7.13	9.67	93.51	25.82
56	147	3.67	13.47	6.67	44.49	24.48
60	155	7.67	58.83	14.67	215.21	112.52
63	149	10.67	113.85	8.67	75.17	92.51
68	152	15.67	245.55	11.67	136.19	182.87
72	160	19.67	386.91	19.67	386.91	386.91
628	1684		1550.67		2500.67	1764.67



Elementos necesarios:

$\bar{x} = \frac{628}{12} = 52.3$ $\bar{y} = \frac{16.48}{12} = 140.33$

$\partial x = \sqrt{\frac{1550.67}{12}} = \sqrt{129.22} = 11.4$ $\partial y = \sqrt{\frac{2500.67}{12}} = \sqrt{208.33} = 14.4$

$\Sigma\, x'y' = 1764.67$

Aplicación:

$$|° = \frac{\frac{1764.67}{12}}{(11.4)\,(14.4)} = \frac{147.06}{164.16} = 0.896 \Rightarrow 80.3\ \%$$

Para determinar esta relación o correlación, se obtiene este indicador. Es importante saber que una es que dependa una variable de la otra, y otra cosa será, que sea causa de la otra, lo que no se concluye con este indicador.

Asociación o correlación no es sinónimo de causalidad.

La correlación siempre estará entre -1 y 1, pudiéndose decir que, mientras más se aproxima al 0 existirá menos o nula correlación.

-1 > |° > 1

Cuando nos aproximamos al < 1 diremos que: "… a mayor de tal variable, menos de la otra…" y, por otro lado, cuando la correlación es positiva, esto es que se aproxima a la unidad positiva, podremos decir que "… a mayor de esta variable, mayor de aquella…".

Se puede interpretar en *%;* en tal caso, se eleva al cuadrado, el coeficiente y después se lleva a % Se diría, en tal caso:

"… en X%, se explican los cambios, por las variaciones de…".

También puede utilizarse la siguiente fórmula:

$$I. = \frac{\sum X \; Y - n \; \bar{X} \; \bar{Y}}{(n-1) \; Sx \; Sy}$$

En donde: $Sx = \sqrt{\dfrac{\sum X^2 - n \; \bar{X}^2}{n-1}}$ $\qquad Sy = \sqrt{\dfrac{\sum Y^2 - n \; \bar{Y}^2}{n-1}}$

ASOCIACION

Es la misma idea de la correlación, con la diferencia que algunos autores la consideran, que cuando los datos son de escala cuantitativa, se usa el índice de correlación; cuando los datos son de escalas cualitativas o combinados con cualitativa y cuantitativa, se usa el Índice de Asociación.

REGRESION

Es la línea de tendencia que sirve para describir un polígono de frecuencias o gráfica de correlación.

Además de la objetividad que nos pueda dar el gráfico que describe a la serie, trazando la línea de tendencia, nos podemos dar cuenta si ésta asciende o desciende en la apreciación general y además nos puede dar la opción de proyectar o calcular, en base a ese comportamiento; es importante que cuando se hagan proyecciones, con esta línea, no se lleve a periodos muy largos.

Cálculos de tendencia

Descripción fenomenológica

Serie cronológica:

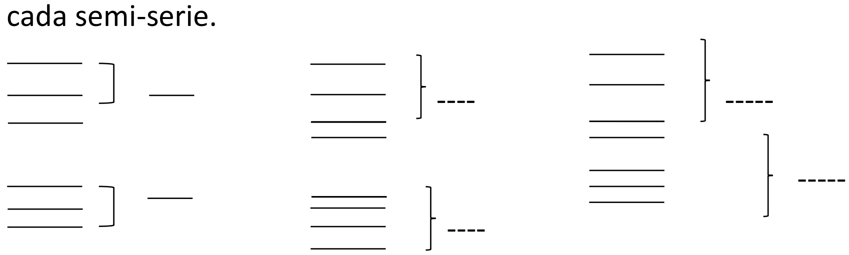
- Secular —> largos periodos
- Cíclicas —> altas y bajas periódicamente
- Estacionales —> durante el año
- Irregulares —> sin patrón gráfico
- Grafía: polígono de frecuencias.
- Métodos de cálculo:
- Mano libre o mano alzada
- Semipromedios
- Promedios móviles
- Mínimos cuadrados
- Por diferencia

Mano libre:

Se traza la línea, a Juicio Individual, para aproximar, gráficamente, una línea, al conjunto de las observaciones.

Semi promedios:

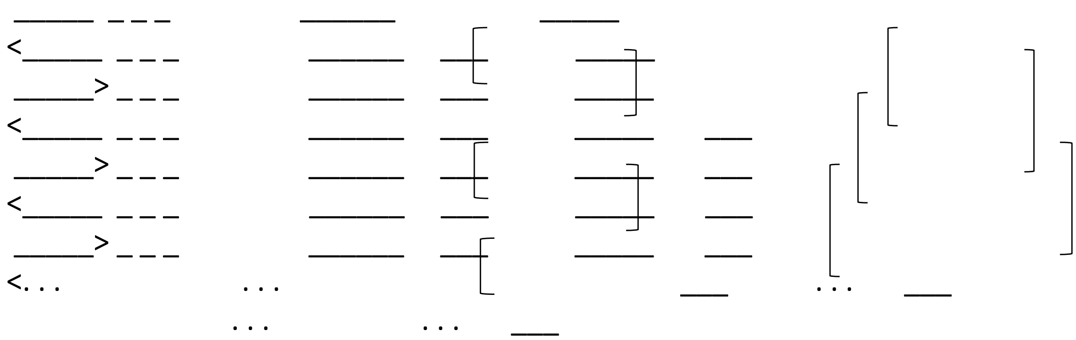
a) Se divide a la serie en dos partes Iguales; si la serie es impar, se considera al valor central, común para ambas partes.
b) Se promedia cada mitad.
c) Se ubica cada punto, en la gráfica, en el lugar que corresponde a la parte central de cada semi-serie.



Se traza la línea, basándose en los puntos resultantes. Cuídese de no "proyectar" más allá de 2 ó 3 períodos.

Promedios móviles: (suavizar variaciones}
Se promedian las observaciones de 2 en 2, de 3 en 3, 4 en 4, 5 en 5, etc...

2 en 2 3 en 3 5 en 5



Para obtener los últimos y primeros, se promedian con los primeros y los últimos.

Para unirse los puntos, se realiza con continuidad; evítese la línea quebrada.

Mínimos cuadrados:

Recta: Y - a + b X

a y b son parámetros, mientras no tienen un valor

son constantes, cuando se determina el valor.

X y Y son variables.

Ecuaciones lineales de la recta:

$$Y = a + b X$$

$$\sum Y = a N + b \sum X$$

$$\sum X Y = a \sum X + b \sum X^2$$

La constante "a" se obtiene con sumatoria (), en ambos miembros; la "b", multiplicando por X y sumándose luego.

Para despejar una literal, se determina el numerador, ignorando los términos de la literal que se despeja.

En el denominador, serán los miembros de ambas incógnitas.

Obtención de:

$$a = \frac{(\sum Y)(\sum X^2) - (\sum X)(\sum XY)}{N \sum X^2 - (\sum X)^2}$$

$$b = \frac{N \sum XY - (\sum X)(\sum Y)}{N \sum X^2 - (\sum X)^2}$$

Por diferencias: (suma y resta)

Se busca la forma de eliminar cualquiera de las incógnitas, recordando que las ecuaciones no se alteran, siempre que se multiplique, divida, sume o reste, en ambos miembros.

EJEMPLO:

Integremos un ejemplo manejando los diferentes métodos.

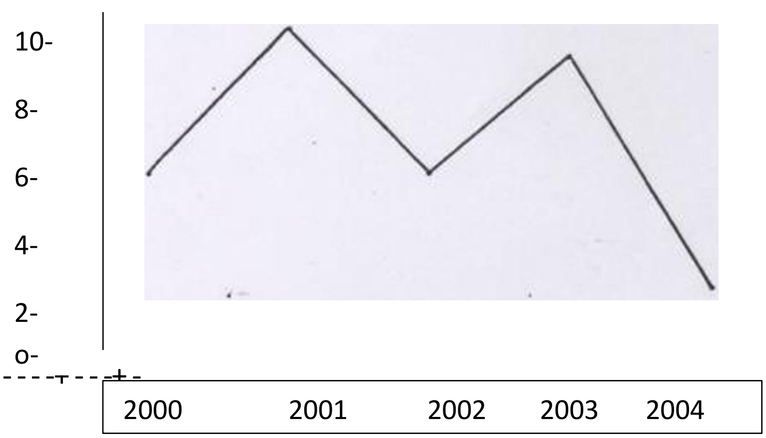
Defunciones por sarampión

Estados Unidos Mexicanos

2000 – 2010

Número (miles)

AÑO	DEFUNC.
2000	6 995
2001	11 891
2002	7 107
2003	11 504
2004	2 609



Mano alzada: Se tira la línea conforme el sentido común y se

Calcule que describe a la gráfica.

Semi promedios:

6 995 + 11 891 + 7 107 =25 993 / 3 = 8 664

7 107 + 11 504 + 2 609 = 21 220 /3 = 7 073

Se coloca el primer punto,a la altura de 8 064 y, en relación a la línea de las X, sobre el final de 1970 que es el año, que está al centro del primer semipromedio. El otro punto se coloca en 1972, a la altura de 7 073, que es el segundo semipromedio. Con estos dos puntos ya se puede trazar la línea de regresión.

Promedios móviles:

6 995
-------- > 18 886 / 2 = 9 443
11 891
-------- > 18 998 / 2 = 9 499
7 107
-------- > 18 611 / 2 = 9 306
11 504
------------- > 14 113 / 2 = 7 056
2 609

Conforme observamos en donde se van a colocar los puntos, en la gráfica, según indican los promedios de cada 2 observaciones, éstos irán al centro de cada año señalado.

Se traza la línea siguiendo a los puntos. En este caso no se debe trazar la línea recta; se irá trazando, con suavidad, siguiendo más o menos el sentido de los puntos.

Mínimos cuadrados:

X	Y	X	X Y
1 2 3 4 5	6 995 11 891 7 107 11 504 2 609	1 4 9 16 25	6 095 23 782 21 321 46 016 13 045
15	40 106	55	111 159

En donde:

$N = 5$

$\sum X = 15$

$\sum Y = 40\ 106$

$\sum X^2 = 55$

$\sum X\ Y = 111\ 159$

$(\sum X^2) = 225$

Despejar

$$a = \frac{(40\ 106)(55) - (15)\ (111\,159)}{(5)\ (55)\ - 225} = \frac{2\,205\ \ 830\ - 1\ \ 667\ \ \ 385}{275\ \ -\ \ 225} = \frac{538\ \ 445}{55} = 10\ \ 769$$

$$a = 10\ \ 769$$

$$b = \frac{(5)\ (111\ \ 159)\ - (15)\ (40\ \ 106)}{(5)\,(55)\ - 225} = \frac{555\ 795 - 601\ 590}{55} = -916$$ b = - 916

Gráfica

X	Y
0	10 769
5	6 189

$$
\begin{aligned}
Y &= a + b \quad X\\
Y &= 10\ 769 + (-916)\,(0)\\
&= 10\ 769\\
Y &= 10\ 769 + (-916)\,(5)\\
&= 10\ 769 + (-4\ 580)\\
&= 6\ 189
\end{aligned}
$$

Se colocan los 2 puntos (1O 769 y 6 189!) en la gráfica, para que con éstos se trace la línea. Cuando X vale O, Y vale 1O 769;y cuando X vale 5, Y vale 6 189.

Por diferencias:

$$
\begin{aligned}
&Y = a + b \quad X\\
&\sum Y = a\ N + b \sum X && \rightarrow \quad 40\ 106 = 5a + 15b\\
&\sum XY = a \sum X + b \sum X^2 && \rightarrow \quad 111\ 159 = 15a + 55b
\end{aligned}
$$

Para despejar cualquier incógnita se debe buscar la manera de que, multiplicando una ecuación, logre igualarse con alguno de los términos, para que se pueda eliminar una incógnita y, quedando una, se despeje sólo una.

En el caso concreto multiplicaremos por 3, para igualar el término 5a con 15a; y por (-l), en la segunda ecuación, quedará nulificada una literal.

Despejar "b" (eliminaremos (”a")

$$
\begin{aligned}
40\ 106 &= 5a + 15b && \rightarrow \quad 120\ 318 = 15a + 45b\\
111\ 159 &= 15a + 55b && \rightarrow \quad -111\ 159 = -15a - 55b\\
& && \qquad\quad \overline{9\ 159 = \qquad\quad -10b}
\end{aligned}
$$

$$b = \frac{9\,159}{-10} = -916$$

Despejar “a” (basta una ecuación)

40106=5a+(15) - (916)
40 106 = 5a + (13 740)
40 106 + 13 740 = 5a

$$a = \frac{53\ 846}{5} = 10\ 769$$

Gráfica

X	Y
0	IO 769
5	6 189

$$Y = a + b\ \ X$$
$$Y = 10\ 769 + (-916)\ (0)$$
$$= 10\ 769$$
$$Y = 10\ 769 + (-916)\ (5)$$
$$= 10\ 769\ + (-4\ \ 580)$$
$$= 6\ 189$$

ÍNDICE ENDÉMICO

Es el cálculo que se hace para predecir el número de casos de enfermedades que se esperan para el periodo que se inicia, en un lugar determinado.

Es la resultante de aplicar una serie de pasos para calcular el número de casos de cierta enfermedad, que se esperan, semana por semana, basándose en el comportamiento de los mismos, según registro de los últimos 5 años. Este índice o canal endémico se utiliza básicamente dentro de las estadísticas de morbilidad o enfermedades; sin embargo, si observamos la manera de obtenerlo, podría servirnos para calcular el comportamiento de otros eventos de diferente naturaleza y que puede servir este mecanismo para su cálculo.

No sólo determina el número de casos; sino que proporciona límites, mínimo y máximo, para determinar, en algún momento, alguna epidemia o éxito.

Se puede obtener mediante la media ($\bar{X}$ y S) o la mediana (Md y Q). Es recomendable que se realice a través de la Media.

Antes de proceder a su obtención, que independientemente del método que se elija, se debe suprimir la información de todo el año en que se haya detectado alguna epidemia, para evitar inconsistencias en el conjunto; y alteración en lo esperado. Entiéndase que una cosa es epidemia y otra, el que se inicie, modifique o incremente un registro, en alguna época; en este caso se deberá considerar la nueva información.

Siempre se deberá tener un mínimo de información de los últimos 5 años.

El índice endémico siempre debiera hacerse por semanas

a) Siempre se debe contar con la información semanal, recuérdese la obligatoriedad, para que con oportunidad se prevean las fechas de vacunación y se establezca, a tiempo, el nivel de producción de los individuos.
b) Con fines epidemiológicos siempre se deberá contar con el número de casos esperados, para que con oportunidad se prevean las fechas de vacunación y se establezca, a tiempo, el nivel de protección de los individuos.
c) Es mucho mejor contar con el índice semanal, que proporciona más detalle, que el mensual.

OBTENCIÓN

Mediana (Md y Q)

1. Se elabora primeramente el cuadro que contenga la información de acuerdo con el siguiente esquema:

SEM	A Ñ O S					Q_1	Md	Q_3
	1983	1884	1985	1986	1987			
1	23	5	12	10	18	7.5	12	20.5 *
2	7	7	1	17	23	4	7	20
3	9	13	20	23	21	11	20	22
4	5	4	12	25	20	4.5	12	22.5
< >								< >
49	12	10	17	21	15	11	15	19
50	3	2	20	29	15	2.5	15	24.5
51	3	6	26	12	13	4.5	12	19.5
52	-	3	19	11	7	1.5	7	15
TOT	408	446	1086	1025	1328			

*Los Q 1 y 3, no son el 2º y 3er valor de la serie (ordenados); sino que es el promedio del 1º más el 2º y el 3º más el 4º, respectivamente, ya que n + ½ = 5 + 1 / 4 = 6 / 4 = 1.5 (lugar). En el caso concreto, el orden de la primera semana será: 5, 10,12, 18 y 23; así los lugares del Q1, Md y Q3, serán 1.5, 3 y 4.5 respectivamente y los valores: 7.5, 12 y 20.5, y así sucesivamente.

2. Se ordenan, en forma ascendente, los casos presentados, cada semana.
3. El número de casos, del año medial (al que divide a la serie en dos partes iguales, se toman como los esperados para casa semana, del año que se calcula el índice endémico (Md).
4. Se obtienen los cuartiles (Q1 y Q3), como calores mínimo y máximo, a esperar por semana.
5. Se gráfica, limitando a lo esperado (Md), con los cuartiles 1 y 3, y con esto se conforma el “canal endémico).

MEDIA ($\overline{X}$, Sy)

Es recomendable obtener el índice endémico con esta medida de resumen, en virtud de que los datos, de los 5 años, por cada semana, con tomados en cuenta para conformar un panorama más confiable, en el promedio que se espera.

1. Se elabora, primeramente, un cuadro que contenga la información de acuerdo con el siguiente esquema:

SEM	A Ñ O S *						$\overline{X}$	ESPERADO	PROM. MOV	LECTURA	$(Y\text{-}Y')^2$	Y´ 2SY	
	83	84	85	86	87	SUMA						-	+
1	23	5	12	10	18	68	13.6	14.9	11.9	13.0	3.61	8.7	17.3
2	7	7	1	17	23	55	11.0	12.1	15.3	15.4	10.89	11.1	19.7
3	9	13	20	23	21	86	17.2	18.9	15.2	15.6	10.89	11.3	19.9
4	5	4	12	25	20	66	13.2	14.5	16.3	16.0	2.25	11.7	20.3
< >													< >
49	12	10	17	21	15	75	15.0	16.5	16.9	16.2	0.09	11.9	20.5
50	3	2	20	29	15	69	13.8	15.1	14.9	13.7	1.96	9.4	18.0
51	3	6	26	12	13	60	12.0	13.2	12.4	12.4	0.64	8.1	16.7
52	-	3	19	11	7	40	8.0	8.8	12.3	12.1	10.89	7.0	16.4
TOT											238.67		

| 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |

*Se han escrito del año de 1983 hasta 1987, sólo para ejemplificar que sean los últimos 5 años, suponiendo que vamos a calcular el índice endémico para 1988.

2. Se registra la información, semanal, de los últimos 5 años, como mínimo. Cuando se haya registrado un año epidémico dentro de estos 5, se deberá eliminar de este procedimiento; en este caso se tomará otro año, para completar los 5 necesarios.
3. Se hace la sumatoria de los casos, de cada semana.
4. Se promedia.
   Si en algún año, se está seguro que no existe información, se promedia entre los años que se tienen; excepto se tenga 0 (cero), que si se cuenta para promediar.
5. Casos esperados para el año, semanal.

Se obtiene de multiplicar el promedio anterior (columna 4) por el incremento poblacional (Ip).

El incremento poblacional (Ip) se obtiene de dividir la población del año al que se calcula el Índice, entre la población del año medial (el año que se encuentra a la mitad de los 5, del estudio); para el casi concreto tomaríamos la población de 1985, que es el que se encuentra a la mitad de la serie.

$$Ip = \frac{\text{Población del año del cálculo}}{\text{Población del año medial}}$$

Ejemplo: (Para 1988, se tomó de 1983 hasta 1987)

$$\frac{\text{Pobl. 1988}}{\text{Pobl. 1985}} \qquad Ip = \frac{856\ 567}{780\ 182} = 1.0979$$

6. Se calcula el promedio móvil (puede ser de 3 en 3)
   Recuérdese que se suman los 3 primeros valores y se obtiene un primer promedio (se coloca a la mitad de los 3, en la columna de promedios móviles); se deja el primer valor y se toman, ahora, el 2º, 3º y 4º valores, se promedia, se coloca a la altura del 3er valor (en la col. de prom. móviles) y en éste mismo orden se continúa. Para obtener el 1º y último valores de esta columna se tomarán el anterior o posterior (último de la serie de lo esperado y primero de la misma, respectivamente)
   Ejemplo:

SEM
1
2
3
4
49
50
51

*

ESPE RADO
14.9
12.1
18.9
14.5

**

8.8 + 14.9 + 12.1 / 3=
14.9 + 12.1 + 18.9 / 3=
12.1 + 18.9 + 14.5 / 3=
18.9 + 14.5 + 15.4 / 3=

PROM. MOVIL
11.9
15.3
15.2
16.3

52	16.5		
TOT	15.1	19.1 + 16.5 + 15.1 / 3 =	16.9
	13.2	16.5 + 15.1 + 13.2 / 3 =	14.9
	8.8	15.1 + 43.2 + 8.8 / 3 =	12.4
		13.2 + 8.8 + 14.9 / 3 =	12.3
		*	
	5		6

7. Se graficarán los datos obtenidos en la columna del promedio móvil, colocando el punto en la gráfica a la mitad de cada semana.
   Se traza la línea uniendo los puntos, suavizando la misma.

   Cortando la línea trazada, entre punto y punto, se leen los promedios justamente en donde se hace el corte. Para esto, en el corte de la línea se ubica y se "hace" una perpendicular que cae sobra la ordenada de la gráfica de lectura, se obtiene con esto "Y".
8. Se obtiene la Desviación estándar (Sy).
   a) Para cada semana, se saca la diferencia de $Y – Y^2$ y se eleva al cuadrado $(Y – Y^2)^2$
   b) Al final de la columna se suman todas las diferenciales cuadriláteras y se divide entre el número de semanas.
   c) Se saca raíz cuadrada y con esto se obtiene: Sy
   d) Se multiplica por 2 la Sy. ^{2}

$$Sy = \sqrt{\frac{(Y - Y')^2}{n}}$$

Ejemplo:

$$Sy = \sqrt{\frac{238.67}{52}} = \sqrt{4.589807692} = 2.14239$$

2SY= (2) (2.14239) = 4.28

9. Se determinan para cada semana, los límites inferior y superior del "CANAL ENDEMICO", sumando y restando a 2, 2Sy.
   $Y' \pm 2Sy$
   Se escriben en su respectiva columna, según se sume o resta cada valor.
10. Se grafica el índice endémico con la siguiente información:
    a) Se colocan puntos con los casos que se esperan (Y), al final de cada semana.
    b) Se colocan, arriba y abaji de Y, los límites inferior y superior. Con estos últimos puntos se trazan sendas líneas que serán las que limiten "el canal".
11. Durante el año calculado, semana a semana, se vierte la información obtenida en el reporte semanal de los casos detectados, del padecimiento en estudio, cuidando a que altura queda la misma; en caso de alguna inconsistencia a que anotación fuera del "canal", se deberá revisar primeramente el reporte, para revisar si hubiera algún error. Confirmada la exacta información, se procede al análisis detallado del fenómeno y toma de decisiones oportunamente.

Hasta aquí se han visto algunos analizadores o conceptos que nos sirven dentro de la estadística descriptiva. A continuación entraremos a considerar algunos estadísticos o indicadores que se manejan en la Estadística Inferencial; considerando a ésta como un valore que se presume o predice, basándose en experiencias previas.

28. ESTADÍSTICO O INDICADOR

Resultante de aplicar una fórmula o analizador.

Manera de nombrar a la cifra resultante de aplicar una fórmula o analizador, con el cual se puede concluir o determinar en qué estado se encuentran los datos de estudio y que se compara, a manera de ubicación, con respecto a un parámetro.

ESTADÍSTICO O INDICADOR

Es el valor "modelo o estándar" que sirve de base para ser comparado con un indicador.
Notamos la diferencia entre parámetro e indicador. Al obtener un valor ya sea aplicando una fórmula o de estudiar una muestra, éste valor "indicador" se compara con otro valor que ya se tiene bien conocido "parámetro" y que sirve de base para que comparándolo con el indicador se determine si es "aceptable" o dentro de los valores aceptables.

Son una serie de elementos que se calculan para apoyarse en el análisis de la información y sobre todo, para la Estadística Inferencial; tales como: Tamaño de la muestra, pruebas de significancia, probabilidad, etc…

MUESTREO

Es la técnica, mediante la cual, se determina un número (n) de sujetos u observaciones para estudiar; en lugar de hacerlo con todo el universo (N).

Se usará el muestreo cuando, entre otros motivos:

El sujeto o elemento se degrade o destruya al haber sido estudiado.

El universo sea infinito o inmenso.

Sea tan semejante que "… para muestra basta un botón"

Para mayor precisión en el estudio, ya que puede ser tan grande o tan vasto, el universo, que por el cansancio o cualquier otro motivo se cometan errores.

CARACTERISTICAS

Depende de:

La variabilidad $(\sigma \;\hat{o}\; p)$

La precisión deseada $(E \;\grave{o}\; (X - \mu))$

El nivel de confianza (Z)

Los recursos disponibles

Condición:

Todos los elementos deben de tener la misma probabilidad de ser elegidos al azar.

TIPOS

Aleatorio simple

Cuando en una tabla de números aleatorios se van seleccionando los elementos a estudiar, o

En una cajita o tómbola se colocan todos los números del universo, y de ahí, se irán tomando los números

Seleccionados para ser estudiados.

Estratificado

a) Posibilidad de obtener valores en cada estrato
b) Se debe contar con un mapa del área y numeradas cada una de las casas.
c) Se combina con el muestreo sistemático seleccionando al azar cualquier:
   - Punto cardinal (N, S, E, O) O
   - Derecha o izquierda D
   - Un número de inicio 17
   - Otro número 6

Ejemplo:(suponiendo que, al azar se tomaron las anteriores decisiones):

Así puedo entrar a la localidad por el Oeste; caminar a la derecha; contar hasta la casa número 17; en ésta. Se Inicia la primera encuesta, y continuar cada 6 casas, levantando las demás encuestas, hasta completar el número deseado, según haya resultado el tamaño de muestra (n).

Por conglomerados

Con el mapa, el cual habrá sido subdividido o estratificado, se numeran éstos y al azar se elige o eligen el número de estratos a estudiar en su totalidad, ya que en este caso el estrato elegido se estudia completamente, y el número personas o familias completarán aproximadamente, el tamaño de muestra deseado.

Sistemático

Con los expedientes, o encuestas, o tarjetas, o casas, se elige cualquier número al azar y con éste se inicia la primera toma de datos; se elige otro número al azar y este número indicará cada cuantos se irá tomando otro elemento a estudiar.

Combinado

Cuando se hace el muestreo con uno o más de los tipos, antes descritos

TAMAÑO DE LA MUESTRA
FORMULAS
Números absolutos $(\overline{X})$:

Cuando se desconoce el universo (N)

$$n = \frac{[z\ \sigma]^2}{E}$$

Cuando se conoce el universo (N)

$$n = \frac{\dfrac{[z^2\ \ \sigma^2]}{E^2}}{1 + \dfrac{1}{N}\ \ \dfrac{[Z^2\ \ \sigma^2]}{E^2} - 1}$$

Números relativos [proporciones (p + q = 1)

Cuando se desconoce el universo (N)

$$n = \frac{z^2\ p\ q}{E^2}$$

Cuando se conoce el universo (N)

$$n = \frac{\dfrac{Z^2 \; q}{E^2 \; p}}{1 \; + \; \dfrac{1}{N}\left[\dfrac{Z^2 \; q}{E^2 \; p} - 1\right]}$$

En donde: z= Nivel de confianza (ej. : 95%, 1.96; 99%, 2.56)

σ =Desviación estándar

E= precisión aceptada (? – µ) (ej.: ± 0.05)

Diferencia que estoy dispuesto a aceptar entre los resultados obtenidos de la muestra, en relación al valor que ya tengo como conocido.

Puede ser el mismo error estándar (σ_x)

p= evento "favorable"

q= evento "desfavorable" (1-p= q)

EJERCICIO

Supóngase que se desea muestrear en una comunidad para levantar una encuesta. Se necesita determinar el tamaño de la muestra.

Condiciones:

Variabilidad.

Conozco, de antemano, algunos valores

Pacientes: σ= 3.3 σ_x =1.5

Homosexualidad: p= 0.1 (10%)

Población, N= 237 458 personas de 15 y más años

Precisión deseada

Determino 0.05 (5% como nivel de significancia,)

Nivel de confianza

Determino 1.96 unidades tipificadas (desviación estándar) para un 90% de confianza

Con estos valores se puede determinar el tamaño de la muestra, aplicando cualquiera de las formulas, puesto que contamos con valores absolutos (x) o relativos (p)

No necesariamente se tienen que calcular diferentes tamaños de muestra para los diferentes valores. Basta calcular uno; éste, sin embargo, tiene que ser con el que represente "mayor" variabilidad, el que tiene mayor % (p) o mayor variación (σ)

Aplicación:

Valores absolutos

Desconociendo el universo

$$n = \left[\frac{Z\sigma}{E}\right]^2$$

$$n = \frac{(1.96)^2(3.3)^2}{(1.5)^2} = \frac{(3.8416)(10.89)}{2.25} = \frac{= 41.835024}{2.25} = 18.593344 = 19 muestras$$

Conociendo el universo

$$n = \frac{\left[\frac{Z^2\sigma^2}{E^2}\right]}{1 + \frac{1}{N}\left[\frac{Z^2\sigma^2}{E^2} - 1\right]}$$

$$n = \frac{\frac{(1.96)^2(3.3)^2}{(1.5)^2}}{1 + \frac{1}{237458}\left[\frac{(1.96)^2(3.3)^2}{(1.5)^2} - 1\right]}$$

$$= \frac{\frac{(3.8416)(10.89)}{2.25}}{1 + 0.00000421127\left[\frac{(3.8416)(10.89)}{2.25} - 1\right]}$$

$$= \frac{41.835024}{1 + 0.00000421127(41.835024 - 1)} =$$

$$= \frac{41.835024}{1 + 0.00000421127(40.835024)} =$$

$$= \frac{41.835024}{1 + 0.00017196731} = \frac{41.835024}{1.00017196731} =$$

$$= 41.8278309804 = \qquad = 42 muestras$$

Valores relativos

Cuando se desconoce el universo (N)

$$n = \frac{Z^2 pq}{E^2}$$

$$n = \frac{(1.96)^2 (0.1)(0.9)}{(0.05)^2} = \frac{(3.8416)(0.1)(0.9)}{0.0025} =$$

$$= \frac{0.345744}{0.0025} = 138.2976 \qquad = 138 muetras$$

Cuando se conoce el universo (N)

$$n = \frac{\frac{Z^2 q}{E^2 p}}{1 + \frac{1}{N}\left[\frac{Z^2 q}{E^2 p} - 1\right]}$$

$$n = \frac{\frac{(1.96)^2 (0.9)}{(0.05)^2 (0.1)}}{1 + \frac{1}{237458}\left[\frac{(1.96)^2 (0.9)}{(0.05)^2 (0.1)} - 1\right]} =$$

$$= \frac{\frac{(3.8416)(0.9)}{(0.0025)(0.1)}}{1 + 0.00000421127\left[\frac{(3.8416)(0.9)}{(0.0025)(0.1)} - 1\right]} =$$

$$= \frac{\frac{3.45744}{0.00025}}{1 + 0.00000421127\left[\frac{3.45744}{0.00025} - 1\right]} =$$

$$= \frac{13829.76}{1 + 0.00000421127(13829.76 - 1)} =$$

$$= \frac{13829.76}{1 + 0.00000421127(13828.76)} = \frac{13.829.76}{1 + 0.05823664212} =$$

$$= \frac{13829.76}{1.05823664212} = 13068.6837419 \qquad = 13069 muestras$$

α GRADO DE SIGNIFICANCIA

Es el valor que se determina como "diferencia significativa" entre lo que se encuentre como valor (es) contra lo que esperábamos encontrar, basados en experiencias conocidas o anteriores. Siempre deberemos considerar un margen de diferencia, en virtud de que no se puede ser tan preciso, sobre todo cuando, se infiere o, calcula algún resultado.

Normalmente es la diferencia de 100 – Z (considerando a Z, como el % de confianza con la que se está trabajando). Así, pues, se trabaja con °Nivel de confianza" y con nivel o "grado de significancia".

Ejemplo: si decidimos darle un 95% (0.95) de confianza a nuestra investigación, entonces el grado de significancia α= 0.05

Nótese que se simboliza, el grado de significancia, con alfa (α).

Evidentemente existen diferencias, pero algunas resultan ser "significativas" y otras no.

Ѵ GRADOS DE LIBERTAD

Es un valor que sirve para que, en base a éste, se encuentre en las tablas de distribución, determinadas para T de Student (t), o Ji cuadrada, etc. En estas tablas se piden los grados de libertad y éstos se obtienen de 2 diferentes maneras, según la prueba de significancia de la que se trate (t ó X^2)

Para t, será el número de elementos (n) menos 1; Ѵ= (n – 1)
Para X^2 , será: columnas – 1, multiplicado, por renglones – 1
Ѵ= (c- 1) (r – 1)

Nótese que se significa con Ѵ ó g.l., los grados de libertad.

Una vez determinado el número de grados de libertad, que en las tablas se encuentran en la columna matriz, se busca en la columna que corresponda a los grados de significancia que se haya determinado y en la coordenada correspondiente en la tabla, se encuentra el "valor máximo que se puede aceptar" para la toma de decisión, en cuanto a si se acepta o rechaza un valor, según haya sido la hipótesis nula o alternativa.

Z NIVEL DE CONFIANZA

Es el valor, en porcentaje o decimales que se determina para indicar con cuanta confianza estamos trabajando. Esta va de 0 a 100, ó de 0 a 1.

Es común que este nivel de confianza se diga como: 80%, 85% , 90%, 95%, es raro que se diga 100%. Puede darse el valor que uno desee.

Si recordamos lo estudiado en la "curva normal" encontraremos que esto es igual a las unidades tipificadas o a la desviación estándar, y que se establece con valores de este tipo; es decir, que para decir que se trabaja con el 95,45 % de confianza, estamos trabajando con 25 (2 desviaciones estándar); si queremos decir que con el 95% sería Z= 1.96; con 90%, Z= 1.64, etc…

EE, σ_X ERROR ESTANDAR

La idea es similar a lo que corresponde a la "desviación estándar"; sin embargo es importante hacer la siguiente reflexión:

Recordamos que cuando vimos la Desviación Estándar obtuvimos como valor, en el ejemplo, S= 3.3, y que en relación a la media= 6.4, se podría decir que 3.3 era el promedio que se diferenciaba cada valor, con respecto a la media. Ahora bien, al obtener en ese mismo ejemplo al error estándar, σx = 1.5 dado por:

$$\sigma_x = \frac{S}{\sqrt{n}}$$

Diremos que todavía, alrededor de ese 3.3 de la "S" podemos esperar un "cierto margen de movimiento" de 1.5 más o menos en relación de 3.3; esto es: que la desviación estándar vale 3.3 pero puede estar, en la Interferencia Estadística, (el valor de 3.3) entre 1.8 y 4.8 (3.3 ±1.5)

Concluyendo: la media vale 6.4

Esta se puede encontrar, con un 68.27% de confianza, entre 3.1 y 9.7 $\longrightarrow$ (6.4 ±3.3)

La desviación estándar vale 3.3
Esta se puede encontrar, entre 1.8 y 4.8

(a+b) ELEMENTOS DEL BINOMIO

Evidentemente éste proceso no debiera corresponder a estos apuntes, puesto que es una mecanización algebraica que debiera ser conocida; sin embargo, dada la importancia y necesidad, como base para la probabilidad, me permito recordarla, en su desarrollo.

Para encontrar el resultado de: $(a+b)^2$ cualquiera recordaría, hasta de memoria, que es: “el cuadrado del primer término, más el doble producto del primer término por el segundo, mas el cuadrado del segundo término”. = $a^2 + 2ab + b^2$

Ciertamente esto resulta difícil; sin embargo, cuál sería el resultado de: $(a+b)^{54}$ ó de cualquier otro exponente?

Veamos con un ejemplo sencillo, como se puede resolver.

$$(a+b)^7 = a^7 + 7a^6b + 21a^5b^2 + 35a^4b^3 + 35a^3b^4 + 21a^2b^5 + 7ab^6 + b^7$$

Con el binomio anterior, vayamos observándolo y deduciendo algunas reglas que nos harán fácil su obtención:

Algunos significados:

Término= cada parte, separada por + ó - en el desarrollo; Ej., a^7 ó $21a^5b^2$ etc

Literal= a ó b, en el caso concreto.

Exponente= potencia a la que se eleva un valor.

Coeficiente= valor que antecede a las literales, en cada término

1._ Los términos serán, uno más, que el exponente

2._ Extremos: izquierda; parte del primer término literal, con exponente igual al que indica el exponente del binomio a resolver; y va disminuyendo en cada término. Del segundo término literal, sucede lo contrario.

Nótese que en todo caso, siempre los exponentes de las literales sumaran, lo mismo que el exponente del binomio.

3._ Los coeficientes se pueden obtener:

A.- Multiplicando el ultimo coeficiente obtenido (partiendo del 1, en el primer término) por el primer exponente de las literales y se divide entre los términos que se llevan.

B.- Para obtener el coeficiente de un determinado término, sin desarrollar el binomio.

Primero se definen las literales, con sus respectivos exponentes.

Se resta, al exponente del binomio, el término deseado y se le agrega 1; éste será el exponente de la primera literal.

a) Al término deseado, se le resta 1, éste será el exponente de la segunda literal.

   En cualquiera de los casos se obtiene el exponente, de la otra literal, completando la suma de las dos literales, para que complete, lo que se indica en el binomio.

Para el coeficiente, se desarrolla: C_{nr} en donde

n= exponente
C= combinación
r= exponente de la segunda literal

Ejemplo: Encontrar el 18° termino de $(a+b)^{27}$

Exponente de las literales:

a) Exponente del binomio= 27; término deseado= 18°; por lo tanto: 27 – 18= 9 + 1= 10; éste es el exponente de "a".
   Para "b" se busca lo que falta para 27; 27 – 10= 17 (éste es el exponente de "b")

b) Coeficiente: $a^{10}b^{17}$

$$C \quad = \frac{27!}{18!(27-18)!} = \frac{27!}{18!9!} =$$

2718

$$= \frac{27x26x25x24x23x22x21x20x19x18!}{9x8x7x6x5x4x3x2x1x18!} = \frac{1700755056000}{362880} =$$

$= 4686825$ así el término 18° de $a^{10}b^{17}$

PROBABILIDAD

La estadística inferencial contempla, básicamente, probabilidad, muestreo y pruebas de significancia, entre otros elementos o procesos.

Necesariamente se requiere información o experiencias anteriores, de los datos que se manejan, para poder prever (inferir) la esperanza matemática de valores que se esperan. Es importante que diferenciemos, que por un lado es el resultado de un porcentaje o tasa o media; por otro lado con qué probabilidad se encontrará en la población que estudiamos o muestremos.

Tratemos, primeramente, de definir qué es probabilidad. Esto lo daremos a través de

$$p = \frac{h}{n}$$

Esto significa que en un evento (n), se pueden dar diferentes resultados o respuestas (h), y la probabilidad está dada por la expresión fraccionaria de la formula anterior (p).

Ejemplos:

1. Al tirar una moneda al aire, sabemos que existe sólo una probabilidad de que resulte águila, de dos posibles respuestas (águila o sol); p= 1/2 =
2. Al tirar un dado y deseamos que salga 3; sabemos que sólo existe un 3, en 6 diferentes respuestas (1, 2, 3, 4,5 ,6), en virtud de que tiene 6 caras. P= 1/ 6 = 0.166666 = 0.17
3. Supongamos que el porcentaje de homosexualidad es de 10; así la probabilidad (de entrada), es=
4. Supongamos que el número de pacientes, promedio mensual, es de 6.4 estamos entendiendo este promedio, como probable de encontrar en alumnos de características similares a la población estudiada =6.4

Nos hemos encontrado, en los ejemplos anteriores, la probabilidad con que se presenta un fenómeno y que éste puede ser con valores relativos o valores absolutos.

Ahora bien, estos valores son los que sirven de base para las operaciones que se requieren en el cálculo de la probabilidad.

Debo señalar que la probabilidad, nunca deberá ser mayor que 1 o menor que 0

Existen diferentes modelos matemáticos que se pueden utilizar para encontrar la probabilidad. Depende del tipo de valor (relativooabsoluto) y del tipo de estudio que se realiza para decidirse por el modelo o distribución de probabilidad adecuada. Podríamos decir que para valores absolutos se usa la Distribución Normal,puesto que en sus elementos requiere de estos valores; para los relativos, se usaría la Binomial puesto que se conoce: cuanto si y no (p=si, q=no); cuanto favorable y desfavorable.

Distribuciones

NORMAL (Gauss)

$$Z = \frac{\bar{x} - \mu}{\partial}$$

BINOMINAL (Bernulli)

$$\mathrm{p} = \quad {}_{n}c_{x}p^{x}\, q^{n-x} = \frac{n!}{x!\,(n-x)!}p^{x}q^{n-x}$$

POISSON

$$p = \frac{\lambda^{x}e^{-\lambda}}{X!}$$

En donde:

p + q = 1 Ej: 0.1 + 0.9 = 1

p= probabilidad de que si

q= probabilidad de que no

Ej. Homosexualidadp= 0.1
No homosexualidad q = 0.9

Algunas propiedades:

Concepto	Normal		Binomial	Poisson
Media	$\bar{x}$	μ	Np	λ
Varianza	S^2	σ^2	Npq	λ
Desv. Estándar	S	σ	$\sqrt{npq}$	$\sqrt{\lambda}$
Error estándar	S_x	σ_x	$\sqrt{\frac{pq}{n}}$	

Para encontrar la Probabilidad, en la mayoría de los casos se deben consultar las tablas de valores para cada distribución. Estas tablas se encontrarán en algún libro de estadística.

Antes de poder aplicar las formulas, en lagunos ejemplos necesitamos recordar cómo encontrar los valores de un área, bajo la curva normal, o valores de Z.

Ya hemos visto la derivación estándar (σ), pues ésta es lo mismo que unidades tipificadas, o "Z". Recordamos que la curva norma, en el centro, parte de 0, hacia la derecha y a la izquierda, que cubre toda el área, y vale 1 ó 100% ahora bien, no sólo es de 6 desviaciones estándar (3 positivas y 3 negativas) en todo el área; sino que entre cada una puede ser infinitecimal. Asi podemos decir: 1σ ó -2σ. ó 1.68σ ó -2.67σ etc…

EJEMPLOS:

Supóngase que en 200 estudiantes, de una universidad, se encontró un promedio ($\bar{x}$), de estatura, de 171 cms, y una desviación estándar (σ) de 18 cms.

1.- ¿Cuántos estudiantes están entre 169 y 174 cms?

2.- ¿Cuántos estudiantes están entre 148 y 185 cms?

3.- ¿Cuantos son más grandes de 175 cms?

4.- ¿Cuántos son menores que 166 cms?

5.- Entre que estaturas se encuentran los estudiantes, cuando el área está limitada entre: Z= -1.63 y z= -084 ?

6.- Entre que estaturas se encuentran los estudiantes, cuando el área está limitada entre: z=-1.96 y z=1.96 ?

Desarrollo:

$$Z = \frac{\bar{x} - \mu}{\sigma}$$

1.-$z = \frac{168.5-171}{18} = -0.13888 \simeq -0.14$

$z = \frac{174.5-171}{18} = 0.19444 \simeq 0.19$

*Recordemos que se deben tomar los límites verdaderos (media unidad anterior y posterior, al límite real).

Busquemos ahora, en las tablas de valores bajo la curva normal (Z), cuánto vale el área de:

a) Z ==> 0 a -0.14 y encontramos: 0.0160
b) Z==> 0 a 0.19 y encontramos: 0.0359
Como, Z esta de 0 a la izquierda y de 0 a la derecha; sumamos 0.0160 + 0.0359 = 0.519

Nota: en los ejemplos usare el siguiente esquema, que aparenta la línea natural de los numeros; no se tome como tal. Imagínese o supóngase que existe una "Campana de Gauss "y los límites que se indican abajo, junto el sombreado, simulan "el área bajo la curva normal".

--|---------|---------|---------|---------|---------|---------|----

-3σ -2σ -σ 0 σ 2σ 3σ

| |

0.0519

Esto nos dice que 0.0519, ó el 5.19% se encuentran entre 169 y 174 cms. Así, encontramos, de los 200 ==>
200 (0.0519) = 10.38. 10 estudiantes, aproximadamente, se encuentran entre estas estaturas.

2.-$z = \frac{147.5-171}{18} = -1.30555 \simeq -1.31$

$$z = \frac{186.5 - 171}{18} = 0.86111 \simeq 0.86$$

Busquemos ahora, en las tablas de valores bajo la curva normal (Z), cuánto vale el área de:

a) Z==> 0 a -1.31 y encontramos: 0.4049
b) Z==> 0 a 0.86 y encontramos 0.3051

Súmanos, ahora, 0.4049 + 0.3051 =0.71

--|---------|---------|---------|---------|---------|---------|----

-3σ -2σ -σ 0 σ 2σ 3σ

|

0.071

Esto es, 0.71 ó 71%, por lo tanto: 200 (0.71) = 142
Esperemos encontrar a 142 estudiantes, aproximadamente, entre 148 y 185 cms de estatura.

3.- ¿Cuántos son más grandes que 175 cms? (mayor de 175 es de 176, en adelante, y 176 inicia desde 175.5) por lo tanto:

$$z = \frac{175.5 - 171}{18} = 0.25$$

Busquemos ahora, en las tablas de valores bajo la curva normal (Z), cuánto vale el área de:

Z==> 0 a 0.25 y encontramos: 0.0987

Como la curva normal vale 1: es decir, del 0 a la derecha =0.5 y 0.5. Del 0 a la derecha =0.5 y 0.5, del 0 a la izquierda, =0.5+0.5 = 1. Entonces nos interesa saber solo de 0.987σ en adelante; como esto se encuentra después del 0 a la derecha; tomamos: 0.5 – 0.987 = 0.4013.

--|---------|---------|---------|---------|---------|---------|----

-3б -2б -б 0 б 2б 3б

| 

0.4013

Esto es, 0.4013 ó 40.13% por lo tanto: 200 (0.4013) = 80.26

Esperamos encontrar a 80 estudiantes, aproximadamente, más altos que 175 cms. de estatura.

4.- ¿Cuántos son menores que 166 cms? (menor que 166 es de 165, para abajo, y 165, para atrás, inicia en 165.5) por lo tanto:

$$z = \frac{165.5 - 171}{18} = -0.30555 = -0.31$$

Busquemos ahora, en las tablas de valores bajo la curva normal (Z), cuánto vale el área de:

Z==> 0 a -0.31 y encontramos: 0.1217

Como nos interesa del -0.31σ para atrás, restamos:

0.5 – 12117 = 0.3783

--|---------|---------|---------|---------|---------|---------|----

-3б -2б -б 0 б 2б 3б

0.3783

Esto es 0.3783 ó 37.83%, por lo tanto 200 (0.3783) = 75.66

Esperamos encontrar a 76 estudiantes, aproximadamente, más bajos que 166 cms de estatura.

5.- Entre qué estaturas se encuentran los estudiantes, ¿cuándo el área está limitada entre; z = -1.63 y z= -0.84?

Ahora el problema se modifica, en este caso lo que si conocemos es Z; en cambio, ignoramos el valor, ahora busquemos $\bar{x}$

$$z = \frac{-\mu}{\sigma} \quad \text{ó} \quad z\sigma = \bar{x} - \mu \quad \text{ó} \quad Z\sigma + \mu = \bar{x}$$

a) $\bar{x}$ = (-1.63)(18)+(171) = 141.66
b) $\bar{x}$= (-084)(18)+(171) =155.88

--|---------|---------|---------|---------|---------|---------|----

117 135 153 171 189 207 225

Los estudiantes están entre 141.7 y 155.9 cms

6.- Entre que estaturas se encuentran los estudiantes, cuando el área está limitada entre; Z= -1.96 y z= 1.96? (casi 2 б)

a) Ẋ = (-1.96)(18)-171 = 135.72

b) Ẋ = (1.96)(18)+(171) = 206.28

--|---------|---------|---------|---------|---------|---------|----

117 135 153 171 189 207 225

Los estudiantes están entre 135,7 y 206.3 cms.

Recordemos que son dos situaciones diferentes:

a) La esperanza matemática de obtener un valor
b) La probabilidad de que se dé, un determinado evento.

Tomando para ejemplificar, un promedio mensual de pacientes, de 6.4 y un 10% de homosexualidad; tratare de explicar lo anterior (notamos que se tiene, valor absoluto y relativo) así pues si en la población que ya conocemos.

Con valores absolutos:

a) Deseo saber cuál es el promedio mensual de pacientes, con 30 alumnos seleccionadas al azar de acuedo con los resultados ya conocidos, es de esperarse que tengan 6.4, en promedio y 3.26 como desviación estándar; esta es la Esperanza matemática.
   Sin embargo se puede manejar esta esperanza, con límites según la confianza o seguridad (Z) con la que se determine.
   Para un 90%

$$\dot{X} \pm 1.64\ 6$$
$$6.4 \pm (1.64)(3.26)$$
$$1.04 < 6.4 < 11.74$$

Esto significa que si bien es de esperarse que los 30 alumnos seleccionados, tengan un promedio de 6.4 se puede expresar que el 90% de las 30 alumnos manifiesta un promedio mensual entre 1 y 12 pacientes.

b) Deseo saber "qué probabilidad" tengo de encontrar a una alumno, de 30 seleccionados al azar, que tenga "exactamente" 3 pacientes, como promedio mensual. Aplico la fórmula de probabilidad "Normal", puesto que son valores absolutos.

$$z = \frac{\dot{X} - \mu}{6}$$

Para que sean 3, exactamente, tomare de 2.5 hasta 3.5 que son los límites verdaderos de 3

$$z = \frac{2.5 - 6.4}{3.26} = \ -1.196319 = -1.20$$

$$z = \frac{3.5 - 6.4}{3.26} = -0.8895705 = -0.89$$

Busquemos ahora, en las tablas de valores bajo la curva norma (Z), cuánto vale el área de:

a) Z ==> O a -1.20 y encontramos: 0.3849
b) Z ==> O a -0.89 y econtramos: 0.3133
c) Restamos, ahora, 0.3849 – 0.3133 = 0.0718 (lo único que se interesa del área)

--|---------|---------|---------|---------|---------|---------|----

-3б -2б -б 0 б 2б 3б

| |

0.0716 ó 7.2%

Esta es la probabilidad que tengo de contar a una alumno que tenga “exactamente” un promedio de 3 pacientes mensuales.

Con valores relativos

a) Deseo saber cuántos son homosexuales, al seleccionar a 30 individuos: Esperanza matemática, NP = (30)(0.1) = 3
   Este es el número que “esperaría” encontrar.
b) Deseo saber “qué probabilidad” tengo de encontrar a 3 individuos homosexuales, de 30, seleccionados al azar. Aplico la fórmula de probabilidad Binomial, puesto que son valores relativos.

p=$nCxpP^xQ^{n-x} = \frac{n!}{x!\,(n-x)!}p^xq^{n-x}$

$$p = 3c3C\ p^3q^{90-9} = \frac{30!}{3!\,(30-3)!}(0.1)^3\ (0.9)^{27}$$

$$= \frac{30!}{3!\,27!}(0.1)^3(o.9)^{27}$$

$$= \frac{30 * 29 * 28 * 27!}{3 + 2 + 1 + 27!}(0.001)(0.581497)$$

$$= 4060(0.001)(0.0581497) = 0.2360879 \simeq 0.24 \text{ ó } 24\%$$

Esta es la probabilidad que tenido de encontrar “exactamente” a 3 individuos homosexuales, de los 30 seleccionados al azar.

Como resulta ser bastante laborioso el obtener la probabilidad, con la distribución Binomial o Pission, existe la posibilidad de aplicar para todos los caso, las distribución Normal, ya que se ajusta debidamente.

Ejemplo: que probabilidad tengo de encontrar de aplicar, para todos los casos, las distribución normal, ya que se sujeta debidamente.

Ejemplo:

Que probabilidad tengo de encontrar de 7 a 11 individuoshomosexuales, de ambos sexos, entre población de 38 personas elegidas al azar? Se sabe por estudios anteriores que la tasa de homosexualidad es de 9.8% (0.098)

De acuerdo a que son valoresrelativos, se debe utilizar la:

Distribución binomial

P(x) = ${}_nC_X$ $p^x q^{n-x}$ = $\frac{n!}{x!(n-x)!}p^x q^{n-x}$

Se debe sacar la probabilidad de 7, de 8, etc… hasta 11

xP(7) = ${}_{98}C_7$ $(0.098)^7$ $(0.902)^{91}$ =

$$= \frac{32x37x36x35x34x33x32x31!}{7x6x5xx4x2x2x1x31!}(0.000000086812553)(0.0408698) =$$

$$\frac{63\ 606\ 030\ 00}{5\ 040}(0.0000000035480129) =$$

$$= 12\ 620\ 25\ 6(0.0000000035480129) =$$

$$p(7) = 0.0447768$$

$$P(8) = 38\ c\ 8\ (0.902)^{30}\ =$$

$$= \frac{38!}{8!\,30!}(0.098)^6(0.902)^{90} =$$

$= \dfrac{38!}{8!\,30!}(0.0000000085076302)(0.453102) =$

$\dfrac{1\ \ 971\ 788\ 000\ 00}{40\ 320}(0.0000000038648256 =$

$= 48\ 903\ 492\ (0.00000000038548256) =$

P(8) = 0. 1 88 51 4

P<9> $=_{98}$ C $_9$ (0.098)^9(0.902)=

$= \dfrac{38!}{9!\,29!}(0.00000000083877446)(0.050233) =$

=163 011 640 (.00000000041881697) =

P(9) = 0.0068272041

P(10)= $_{38}$C $_{10}$(0.90)^{10} (0.902)^{28} =

$\dfrac{83!}{10!\,28!}(0.000000000081707281])\ (0.0556907) =$

$\dfrac{59\ 153\ 664\ 000\ 000}{362\ 880} = (0.000000000041881697 =$

P(9) = 0.0068272041

P(10) = $_{38}$c$_{10}$(0.098)^{10} (0.902)^{28} =

$\dfrac{38!}{10!\,28!}(0.098)^{10}(0.902)^{28} =$

$= \dfrac{38!}{10!\,28!}(0.098)^{10}(0.902)^{28} =$

$\dfrac{38!}{10!\,28!}(0.00000000081707281)(0.0556907) =$

$\dfrac{1\ 715\ 456\ 000\ 000}{3\ 628\ 800}(0.00000000045503396) =$

P(10) = 0.0021510991

P(11) = $_{38}$ C $_{11}$ (0.098)^{11} (0.902)^{27} =

$\dfrac{38!}{11!\,27!}(.098)^{11}(O.902)^{27} =$

$\frac{38!}{11!\,37!}0.0000000000080073135)(0.0617414) =$

$= \frac{48\ 032\ 775\ 000\ 000\ 000}{39\ 916\ 800}(0.0000000000004938279) =$

P(11) = 0-00059490183

Al haber obtenido las probabilidades de cada número (del 7 al 11), se suman las mismas.

p (7) = 0.0447768

p (8) = 0.0188514

p (9) = 0.0068272041

p (10) = 0.0021510991

p (11) = 0.00059490183

SUMA = <u>0.0732014</u>

las probabilidad de encontrar de 7 a 11 individuos homosexuales, es de:

Nota: Existían tablas de valores para la Distribución binal, para localizar la probabilidad, según el tamaño de la muestra.

Distribución de poisson

$p = \frac{\aleph^x e^{-\aleph}}{X!}$puede ser: $\frac{\aleph^x}{X!\,e^{\aleph}}$

Se platea el reciproco de $e^{-\aleph}$ por $\frac{\aleph^x}{X!e^{\aleph}}$

En donde $\aleph = Np = 28\ (0.098) = 3.724$

$e = 271\ 828$

X1 = El valor que se valla necesitando

P(7-11) = $\frac{\aleph^7 e^{-9.724}}{7!}$+ . . . . + $\frac{\aleph^{11} e^{-3.724}}{11!}$

P(7) = $\frac{(3.724)^7(2.271828)^{-3.724}}{7!} =$

$$=\frac{(3.724)^{7}}{7!(2.71.828)^{3.724}} \quad = \quad \frac{9.932.7089}{(5\,040)(41.429782)} =$$

$$= \frac{(9\,932.7089}{208\,806.1} = \qquad 0.047569$$

$$P\,(8)\,\frac{((3.724)^{8}(2.71828)^{-3.724}}{8!} =$$

$$= \frac{(3.724)^{8}}{8!\,(2.7188)^{8.724}} \qquad = = \frac{36.989.408}{(40\,320)(41.429782)} =$$

$$= \frac{36\,989.408}{1\,670\,448.8} = \qquad 0.0221433$$

$$p = (9)\,\frac{(3.724)^{9}\,(2.71828)^{-9.724}}{9!} \qquad =$$

$$= \frac{(3.724)^{9}}{9!\,(2.71828)^{9.724}} \qquad = \frac{137\,748.55}{(362\,880)(41.429782)} =$$

$$\frac{137\,748.55}{15\,034\,039} \qquad 0.0091624447$$

$$p(10) = \frac{(3.724)^{10}(2.71828)^{-9.724}}{10!} =$$

$$= \frac{(3.724)^{10}}{10!} = \qquad = \frac{512\,975.62}{(3.28\,880)(41.429782)} =$$

$$= \frac{512\,975.62}{150\,340\,390} = \qquad 0.0034120944$$

$$p(11) = \frac{(3.724)^{11}(2.71828^{-9.724}}{11!}$$

$$= \frac{(3.724)^{11}}{11!\,(2.71828)^{3.724}} = \qquad = \frac{1.910\ 321\ .2}{(39\ 916\ 899)(41{,}429782)}$$

$$= \frac{1\ 910\ 321.2}{1\ 653\ 744\ 300} \qquad = \qquad 0.0034120944$$

Al haber obtenido las probabilidades de casa número (del 7 al 11) se suma las sismas, vale la pena ir comparando con los resultados, según la Binomial.

Binomial:

P (7) = 0.447768

P (8) = 0.0188514

P (9) = 0.006827041

P (10) = 0.0021510991

P(11) = 0.0059590183

SUMA = $\frac{0.0732014}{0.073}$

Poission

0.47569

0.0221433

0.0091624447

0.0034120944

0.0011551491

$\frac{0.0833319}{0.083}$

Como podemos observar: tanto para binominal, como para la de Possion, resulta muy largo el proceso. Se puede obtener, con bastante similitud, la probabilidad, usando la Distribución.

Distribucion Normal

$$Z = \frac{\chi - \mu}{\sigma}$$

En donde χ varía, de 7 a 11

M Np = 38 (0.098) = 3.724

σ = $\sqrt{Npq} = \sqrt{38(0.098)} = (0.902)$

$\sqrt{3.359048}$= 1.8327706

$$z = \frac{6.5 - 3.724}{1.8327706} = 1.55146647 = 1.51$$

$$z = \frac{11.5 - 3.724}{1.83277706} = 4.242769 = 4.24$$

Busquemos ahora, en las tablas de valores bajo la curva normal (z), cuánto vale el área de:

a) Z ==> 0 a 1.51 σ y encontramos: 0.4345
b) Z ==> 0 4.24 σ y encontramos: 0.5000
c) Restamos, ahora, 0.5000 (lo único que me interesa del área)

--|---------|---------|---------|---------|---------|---------|----

-3б -2б -б 0 б 2б 3б

|

0.0655 ó 6.6%

Comparando:		
	Distribución Binomial	0.073
	Distribución de Poisson	0.083
	Distribución Normal	0.066

Observamos que realmente no hay tanta diferencia

PRUEBAS DE SIGNIFICANCIA

Son los procesos que se realizan para determinar si entre un valor dado, y otro, en la misma o diferentes poblaciones, "difieren significativamente", o que si bien son diferentes esta no es tan importante que se puedan suponer razones de paso o que se está hablando de otra cosa o población distinta.

Podemos decir, como ya se observe en el capítulo anterior, que al existir límites entre los cuales se espera encontrar un valor y si este se encontrara fuera de los mismo, se puede rechazar el valor como correspondiera a la población.

Que se trata, o decir que si existe diferencia significativa.

Para muestras grandes (algunos autores dicen que mayor de 30), se debe hacer esta prueba con la "curva normal" o "Z". Si se trata de muestras pequeñas (<30) se hace la prueba, con Z con x^2 (ji cuadrada) o con t (t de Student).

Ho HIPOTESIS NULA

Es el enunciado que se hace, con base a la (s) hipótesis establecida(s) en la investigación que se realiza.

Ejemplo: En un principio establecí en la investigación, que no había diferencia entre el uso de hormonas o DIU, en cuanto a la presencia o no de leucorrea. Así, establezco:

Ho: Hormonas = DIU

Otro: Espero encontrar como promedio, 6.4

Ho: $\mu = 6.4$

43.- H_1 HIPOTESIS ALTERNATIVA

Es la misma idea que la anterior, difiere, ahora en que su enunciado será:

H_1: Hormonas ≠ 6.4

44.- TIPOS DE ERROR

Una vez que se establece una hipótesis y se hace el proceso de buscar los valores; al resultar éstos nos vemos en la necesidad de decidir si se acepta o no el resultado. Al compararse co las tablas de decisión, si difiere, no hay problema, simplemente.

x^2 JI CUADRADA

Esta prueba se aplica para valores de escala cualitativa o de escala cuantitativa, o combinados.

FORMULA

$$x^2 = \Sigma + \frac{CO - EO^2}{E}$$

En donde:

O= valores observados

E= valores esperados o teóricos

Se necesitan además:

V= grados de libertad = (c-1) (r-1)

C= columna

r= renglones

a= nivel de significancia

Z= nivel de confianza

EJEMPLOS

A) se considera que no existe diferencia significativa entre el uso de hormonas y del DIU (dispositivo intraurerino), como método de control de planificación familiar; en cuanto, a la presencia o no de leucorrea. Se estudiaron a 100 usuarias activas, en dichos métodos, y se encontraron los siguientes datos:

MÉTODO	LEUCORREA SI / NO	TOTAL
DIU	59/56 / 8/11	67
HORMONAL	25/28 / 8/5	33
TOTAL	84 / 16	100

Para calcular los valores esperados (los que se encuentran en letra pequeña) se hace con la siguiente formula:

E= $\frac{parcial}{total}$ x subtotal del grupo

$\frac{67}{100}$ 84= 56; $\frac{67}{100}$ 16 = 11; $\frac{33}{100}$ 84= 28; $\frac{33}{100}$ 16 =5

Desarrollo

$$x^2 = \frac{(59-56)^{\quad 2}}{56} + \frac{(8-11)^2}{11} + \frac{(25-28)^2}{28} + \frac{(8-5)^2}{5}$$

=9/56+9/11+9/28+9/5= 0.818+0.321+1.8=3.1

Grados de libertad (v):

En columnas, 2 modalidades, 2 columnas.

Renglones, 2 modalidades, 2 renglones.

(c-1) (r-1); (2-1) (2-1)= 1

Nivel de significancia: se determina, α=0.05

Decisión

Ho: DIU=Hormonales.; α=0.05; v=1; Z=95%

Se busca en la tabla de valores de x^2

x^2 V=1; α=0.05 = 3.84146

Ahora se toma la decisión de aceptar o rechazar la hipótesis:

Se acepta la Ho si $x^2 \leq$ 3.84146

Ante de la comparación, se acepta la Ho puesto que 3.1 < 3.84146

Se concluye con que no existe diferencia significativa; por lo que se puede decir, que no existe diferencia entre el uso de DIU i Hormonales, en relación a la presencia o no de leucorrea, en esa población estudia.

B)Se desea saber si existe diferencia significativa, entre 11 individuos estudiados, en los cuales se obtiene el promedio de tensión arterial, para cada uno, y se distribuyen de la siguiente manera:

PROMEDIO	O-E	$(O-E)^2$	$(O-E)^2/E$
118	-9	81	0.6377952
120	-7	49	0.3858267
110	-17	289	2.2755905
132	5	25	0.1968503
153	26	676	5.3228346
130	3	9	0.0708661
135	8	64	0.5039370
104	-23	529	4.1653543
132	5	25	0.1968503
138	11	121	0.9527559
125	-2	4	0.0314960
SUMA			14.740155

Aquí el valor esperado es el promedio de los valores ($\dot{X}$)

Se busca en la tabla de valores de x^2

Decisión: V=10; α=0.05 = 18.3070

Ahora se toma la decisión de aceptar o rechazar la hipótesis:

Se acepta la Ho si $x^2 \leq$ 18.3070

Ante la comparación, se acepta la Ho puesto que 14.7 < 18.3070

t T DE STUDENT (gosset)

Así como dijimos que para pruebas de significancia, en pequeñas muestras (n<30), se utiliza X^2; lo mismo sucede con la prueba de t de Student.

Con "t", se puede determinar entre qué y que valores se "espera" el resultado que se estudie; lo mismo que se puede apreciar si un valor difiere "significativamente" de otro; aun en valores de 2 poblaciones diferentes.

La fórmula es similar a la de "Z". Aquí debemos considerar a la curva que, obviamente, ya no es normal; si no semejante, y que las "colas" de la curva, se tomaran en cuenta. Por ejemplo, al tener α= 0.05 y Z= 0.95; se tomara, el 0.05 de signicancia, como 0.025 para una y otra cola.

Existe tabla, propia, de valores para la distribución de t de Student.

Como las formulas van variando, conforme sea la necesidad a estudiar, las iremos estableciendo en cada ejemplo que se plantea.

EJEMPLOS:

Supóngase con el ejemplo que hemos venido manejando; en relación al promedio de pacientes mensual (μ=6.4; α=3.26; n=15), que:

a) deseo conocer entre que valores se encuentra la $\mathcal{X}$; con un 90% de confianza (α=0.05)

b) encuentro que, en una muestra de una población, de estudio, se da un promedio ($\mathcal{X}$) de 5.8 y S=2.9. Difiere significativamente.

c) en la población B se encontró: $\mathcal{X}$ =5.7; S= 2.8 y n= 14. Difiere significativamente de nuestra población de estudio, que arrojo: $\mathcal{X}$ = 6.4; S=3.26 y n=15?

Desarrollaremos, ahora los anteriores incisos.

DESARROLLO:

a) intervalos de confianza

Z=95% α=0.25 n=15 μ= 6.4 α= 3.26 v= 14

Se determinan, primero, los grados de libertad:

V= n-1=15-1 =14

Par el 95%, considerando la significancia de 0.025; se busca en la tabla de t, en el renglón correspondiente a v= 14 y en la columna de 0.975

$$tv = 14, 0.075 = 2.14$$

$\mathcal{X} \pm tz\,(\frac{S}{\backslash\sqrt{\ \ n-1}})$

6.4 ± $6.4\ \pm\ t\ 0.075\ (\frac{3.26}{\sqrt{15-1}})$

6.4 $\pm 2.14\ (\frac{3.26}{\sqrt{15-1}})$

6.4 $\pm 2.14\ (\frac{3.26}{\sqrt{14}}\ )$

$$6.4\ \pm 2.14\ \left(\frac{3.26}{3.7416574}\right)$$

$$6.4\ \pm 2.14\ (0.8712716)$$

6.4 6.4 ± 1.8645213

4.54<μμ8.26

La media, con un 95% de confianza, se espera encontrar entre 4.5 y 8.3

b) prueba de significancia

Se encuentra, en una muestra de 12 alumnos: $\mathcal{X} = 5.8; S = 2.9$

Valores de la población conocida: μ=6.4; α= 3.26

Nivel de significancia: α= 0.025; Z=95%

Grados de libertad: v= n-1 = 12-1 = 11

Ho: μ= 6.4, no existe diferencia significativa

H :1μ≠ 6.4, $si\ existe\ diferencia\ significativa$

Para $tv = 11; 0.075 = 2.20$

Se acepta la Ho, si t se encuentra entre -2.20 a 2.20

Se rechaza la Ho, si t se encuentra fuera de -2.20 a 2.20

$$t = \frac{\mathcal{X} - \mu}{S}\sqrt{n-1}$$

$$t = \frac{5.8 - 6.4}{2.9}\sqrt{12-1}$$

$$t = \frac{5.8 - 6.4}{2.9}\ (3.3166248)$$

$$t = \frac{0.6}{2.9}\ (3.3166248)$$

$$t = \frac{1.9899749}{2.9} = 0.6861982$$

$$t = 0.69$$

Se acepta la HO, es decir, que probablemente, no existe diferencia significativa entre los resultados encontrados en la muestra, de 12 alumnos; comparando con los resultados de la población que ya se conoce; con un nivel de confianza del 95%, puesto que:

$$-2.20 < 0.69 < 2.20$$

c) comparación de 2 poblaciones

A: $\chi 1 = 6.4$, $S1 = 3.26$; $n = 15$

B: $\chi 2 = 5.7$, $S2 = 32.8$; $n = 14$

Con α= 0.05; $v = n1 + n2 - 2 = 15 - 14 + 2 = 27$

$$Ho: \mu 1 = \mu 2$$

$$1o: \mu 1 \neq \mu 2$$

Para decidir

Para $tu = 27; 0.975\ \ 2.05$

Se acepta la Ho, si t se encuentra entre -2.05 a 2.05

Se rechazara la Ho, si t se encuentra entre -2.05 a 2.5

$t = \frac{x_1 - x_2}{\sigma \sqrt{\left(\frac{1}{n1} + \frac{1}{n2}\right)}}$ En donde

$$\sigma = \sqrt{\frac{n1S^2 + n2S_2^2}{n1 + n2 - 2}}$$

$$\sqrt{\frac{(15)(3.26)^2 + (14)(2.8)^2}{15 + 14 - 2}}$$

$$\sqrt{\frac{(15)(10.6276) + (14)(7.84)}{27}}$$

$$\frac{159.414 + 109.76}{27} = \frac{269.174}{27} = 9.9694074$$

$$t = \frac{6.4 - 5.7}{9.97\sqrt{\left(\frac{1}{15} + \frac{1}{14}\right)}} = \frac{0.7}{9.97\sqrt{0.06667 + 0.07143}}$$

$$\frac{0.7}{9.97\sqrt{0.1380945}} = \frac{0.7}{9.97(0.3716107)}$$

$= \frac{0.7}{3.7049595} = 0.1889359$

$$t = 0.189$$

Se acepta la Ho. Es decir, que probablemente no existe diferencia significativa, entre los resultados de las 2 poblaciones.

$$-2.05 < 0.189 < 2.05$$

Con lo anterior doy por terminados estos apuntes que, sinceramente, deseo te sean de utilidad. Espero haberte dado el mensaje de que la estadística, verdaderamente, no es imposible o difícil de manejar, cuando uno desea hacer alguna investigación.

Evidentemente, no es esto la estadística completa, quedan muchas cosas, por fuera, sobre todo en la estadística inferencial. Que, por otro lado, cuando alguien necesita más elementos para un trabajo, desde luego, requiere profundizar estos conocimientos.

Bibliografía

ALEA, V. et al. (1999) Estadística Aplicada a les Ciències Econòmiques i Socials. Barcelona: Edicions McGraw-Hill EUB.

CANAVOS, G. (1988) Probabilidad y Estadística. Aplicaciones y Métodos. México: McGraw-Hill.

DURA PEIRó, J. M. y LóPEZ CUñAT, J.M. (1992) Fundamentos de Estadística. Estadística Descriptiva y Modelos Probabilísticos para la Inferencia. Madrid: Ariel Editorial.

ESCUDER, R. y SANTIAGO, J. (1995) Estadística aplicada. Economía y Ciencias Sociales. Valencia: Tirant lo Blanch.

FERNáNDEZ CUESTA, C., y FUENTES GARCíA, F. (1995) Curso de Estadística Descriptiva. Teoría y Práctica. Madrid: Ariel.

FREEDMAN, D., et al. (1991) Estadística. Barcelona: A.Bosch Ed.

FREEDMAN, D., et al. (1991) Estadística. Barcelona: A.Bosch Ed.

FREIXA, M., et al. (1992) Análisis exploratorio de datos: Nuevas técnicas estadísticas. Barcelona: PPU.

GUJARATI, D. (1997) Econometría Básica. Bogotá: McGraw-Hill.

KMENTA, J (1980) Elementos de Econometría. Barcelona: Vicens Universidad.

MARTíN PLIEGO, F. (1994) Introducción a la Estadística Económica y Empresarial. (Teoría y Práctica) Madrid: AC.

MARTíN PLIEGO, F. y RUIZ-MAYA, L. (1995) Estadística I: Probabilidad. Madrid: AC.

MARTíN PLIEGO, F. y RUIZ-MAYA, L. (1995) Estadística II: Inferencia. Madrid: AC.

MARTíN-GUZMáN, P. y MARTíN PLIEGO, F. (1985) Curso Básico de Estadística Económica. Madrid: AC.

MENDENHALL, W., et al. (1994) Estadística Matemática con Aplicaciones. México: Grupo Editorial Iberoamérica.

MONTIEL, A.M., RIUS, F. y BARóN, F.J. (1997) Elementos Básicos de Estadística Económica y Empresarial. Madrid: Prentice Hall.

NEWBOLD, P. (1996) Estadística para los negocios y la economía. Madrid: Prentice Hall.

PEñA, D. y ROMO, J. (1997) Introducción a la Estadística para las ciencias sociales. Madrid: McGraw-Hill/Interamericana de España.

PéREZ, C. (1995) Análisis Estadístico con Statgraphics. Técnicas Básicas. Madrid: Ra-Ma.

TANUR, J. (1992) La Estadística, una Guía de lo Desconocido. Madrid: Alianza Editorial.

URIEL, E. y MUñIZ, M. (1988) Estadística Económica y Empresarial. Teoría y ejercicios. Madrid: AC.

URIEL, E. y PEIRó, A. (2000) Introducción al análisis de series temporales. Madrid: AC.